Network Topology Uncovers Function, Disease, and Phylogeny

Nataša Pržulj

Department of Computing
Imperial College London

September, 2010

Overview

1. Introduction and Background:
   - PPI and other networks
   - Network analysis and modeling challenges
2. New Network Analysis and Modeling:

   (A) Analysis:
   - Structure *vs.* biological function and disease
   - Network alignment

   (B) Modeling – Geometric Graphs:
   - Graphlet Degree Distributions
   - Network embedding
   - Trained Geometric Model
   - Application: De-noising PPI networks

   (C) Why PPI networks might be geometric?
3. Conclusions

Nataša Pržulj
natasha@imperial.ac.uk

Overview

1. Introduction and Background:
   - PPI and other networks
   - Network analysis and modeling challenges
2. New Network Analysis and Modeling:

   (A) Analysis:
   - ➢ Structure *vs.* biological function and disease
   - ➢ Network alignment

   (B) Modeling – Geometric Graphs:
   - ➢ Graphlet Degree Distributions
   - ➢ Network embedding
   - ➢ Trained Geometric Model
   - ➢ Application: De-noising PPI networks

   (C) Why PPI networks might be geometric?
3. Conclusions

Nataša Pržulj
natasha@imperial.ac.uk

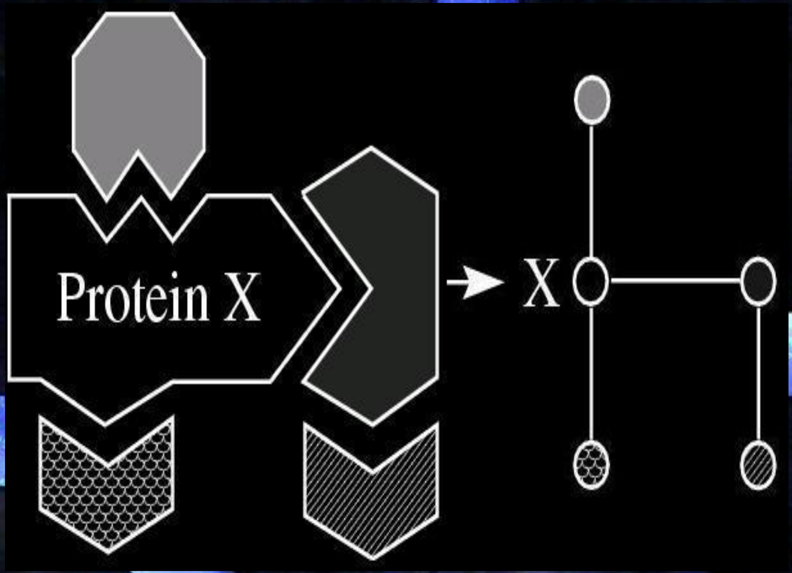
1. Introduction and Background

- Large Networks model many real-world phenomena
  - technological: www, internet, electric circuits,...
  - social: friendship, collaboration, disease spread,...
  - biological:
    - protein structure,
    - transcriptional regulation,
    - metabolic,
    - protein-protein interaction (PPI),
    - …

Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

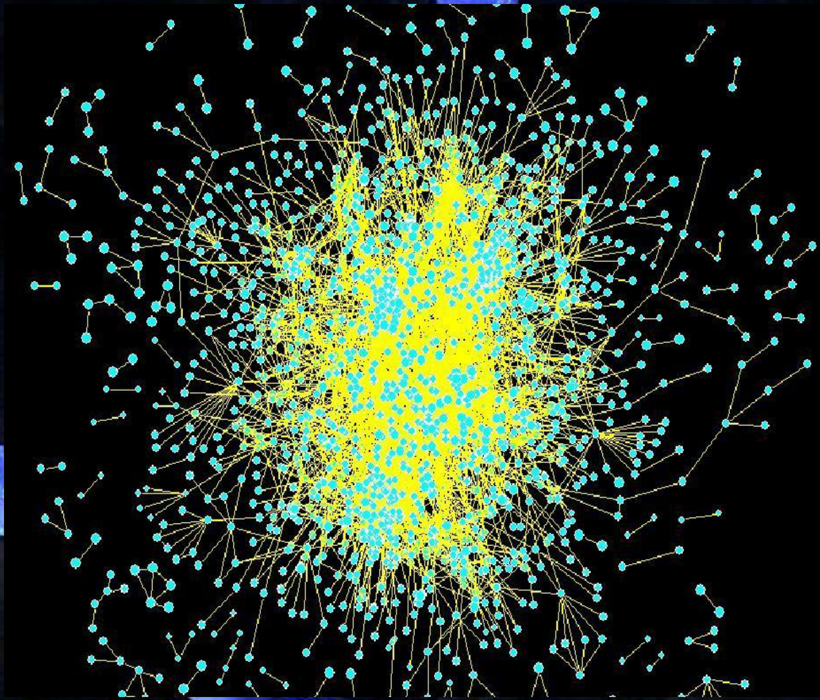
- Large Networks model many real-world phenomena
  - technological: www, internet, electric circuits,...
  - social: friendship, collaboration, disease spread,...
  - biological:
    - protein structure,
    - transcriptional regulation,
    - metabolic,
    - protein-protein interaction (PPI),
    - …



Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

- Large Networks model many real-world phenomena
  - technological: www, internet, electric circuits,...
  - social: friendship, collaboration, disease spread,...
  - biological:
    - protein structure,
    - transcriptional regulation,
    - metabolic,
    - protein-protein interaction (PPI),
    - …



Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

- Large-scale networks in bioinformatics:
  - Technological advances in experimental biology
    - ➢ data
  - Important computational problems
  - Algorithmic and modeling advances contribute:
    - biological understanding (function, disease, pathogens,…)
    - therapeutics
  - ➔ Booming research area

Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

Problems:

1. Noise → revise models as data sets evolve
2. "Hardness" of graph theoretic problems

   E.g. NP-completeness of subgraph isomorphism

   ➔ Cannot exactly compare/align networks

   → heuristics (approximate solutions)

   ➔ Exact comparison inappropriate in biology

   → due to biological variation

Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

Properties of Large Networks (heuristic comparisons)

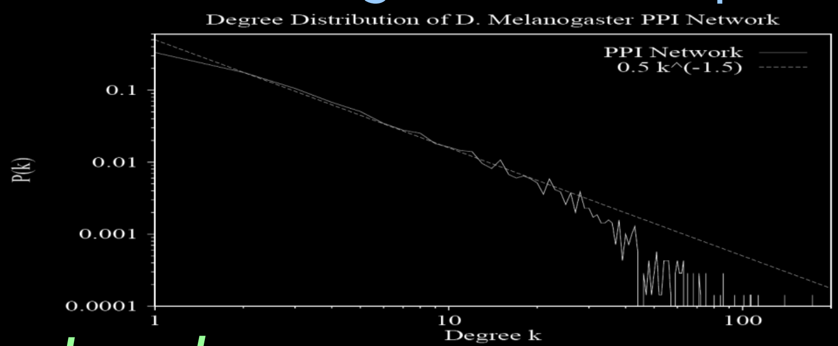
- *Global*
  - Degree distribution
  - Diameter
  - Clustering coefficient/spectrum
- *Local:*
  - network "motifs" and subgraphs (U. Alon's group, '02-'04)

Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

Properties of Large Networks (heuristic comparisons)

- *Global*
  - Degree distribution
  - Diameter
  - Clustering coefficient/spectrum



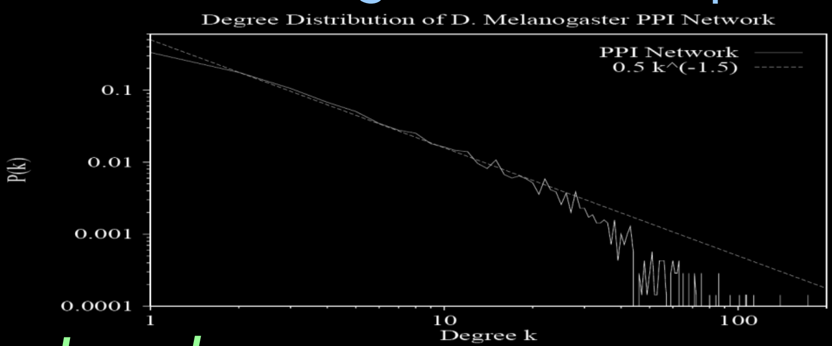
- *Local:*
  - network "motifs" and subgraphs (U. Alon's group, '02-'04)

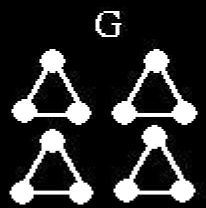
Nataša Pržulj
natasha@imperial.ac.uk

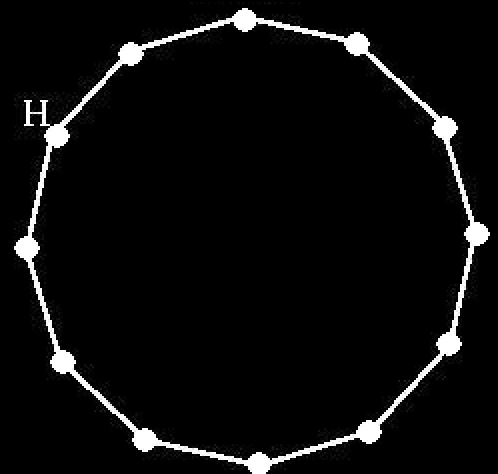
1. Introduction and Background

Properties of Large Networks (heuristic comparisons)

- *Global*
  - Degree distribution
  - Diameter
  - Clustering coefficient/spectrum







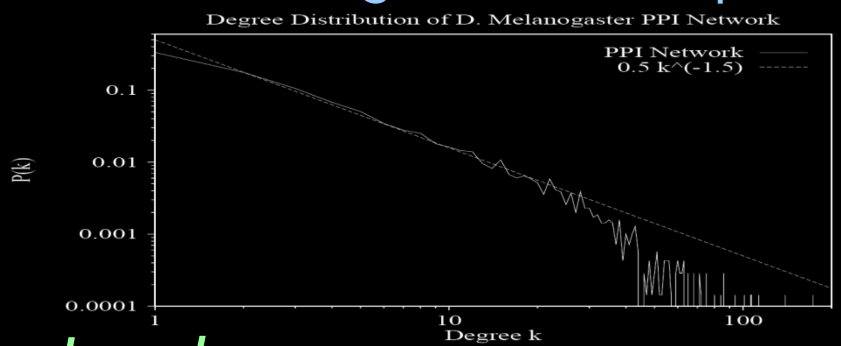
- *Local:*
  - network "motifs" and subgraphs (U. Alon's group, '02-'04)

Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

Properties of Large Networks (heuristic comparisons)

- *Global*
  - Degree distribution
  - Diameter
  - Clustering coefficient/spectrum



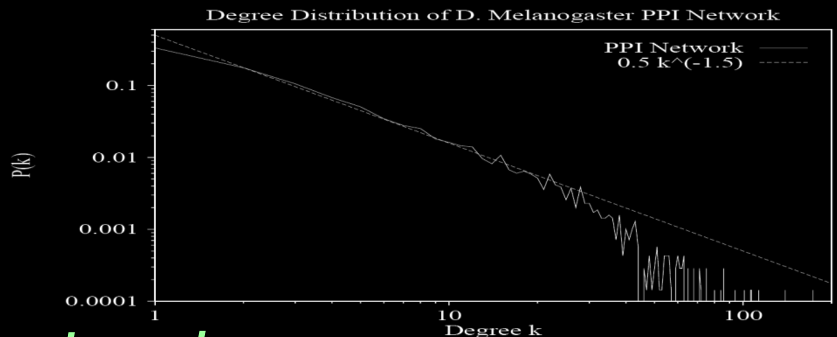
- *Local:*
  - network "motifs" and subgraphs (U. Alon's group, '02-'04)

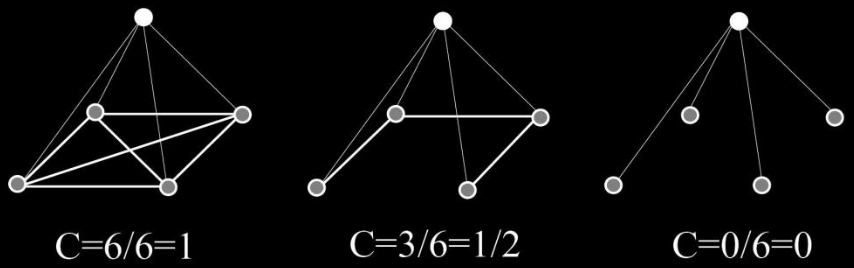
Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

Properties of Large Networks (heuristic comparisons)

- *Global*
  - Degree distribution
  - Diameter
  - Clustering coefficient/spectrum





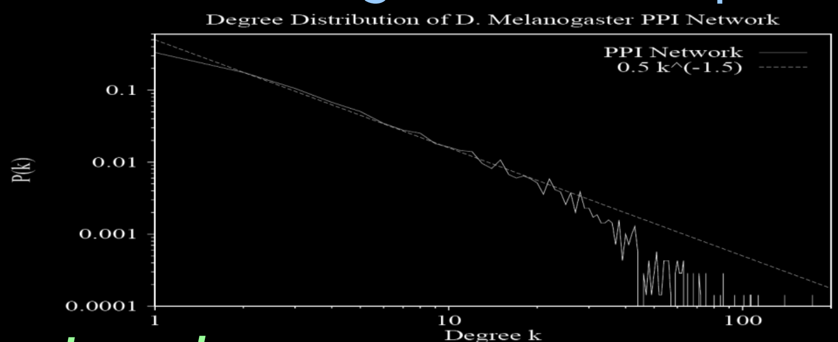
- *Local:*
  - network "motifs" and subgraphs (U. Alon's group, '02-'04)

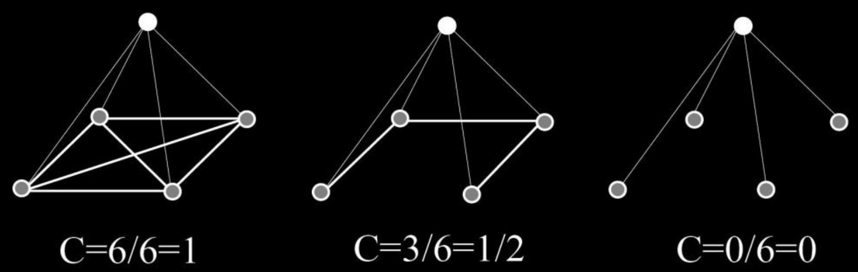
Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

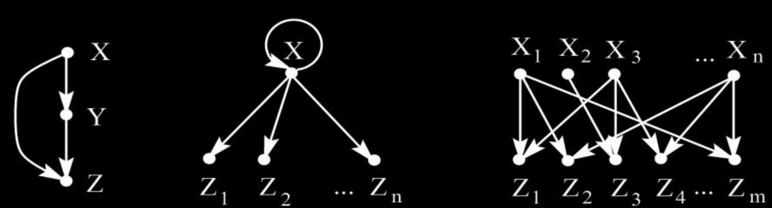
Properties of Large Networks (heuristic comparisons)

- *Global*
  - Degree distribution
  - Diameter
  - Clustering coefficient/spectrum





- *Local:*
  - network "motifs" and subgraphs (U. Alon's group, '02-'04)



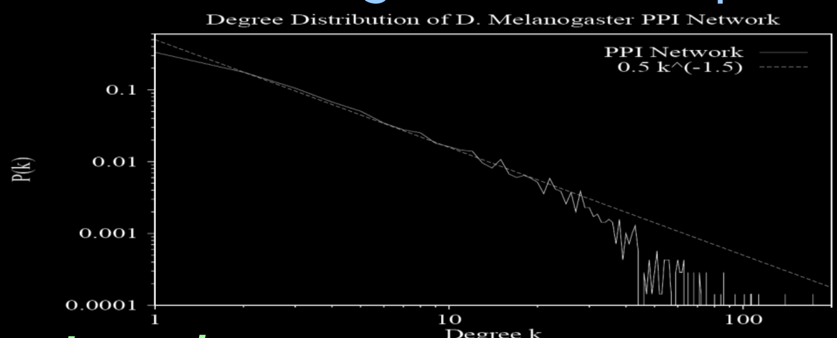
(a) (b) (c)

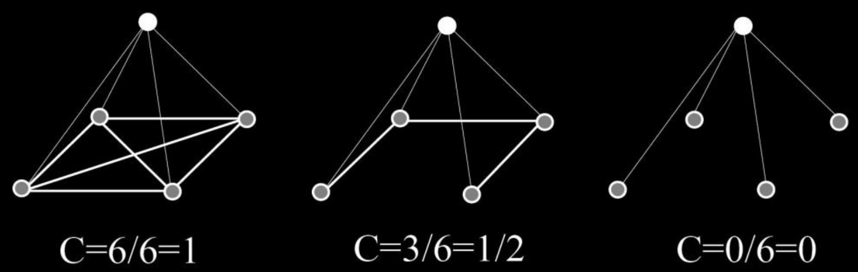
Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

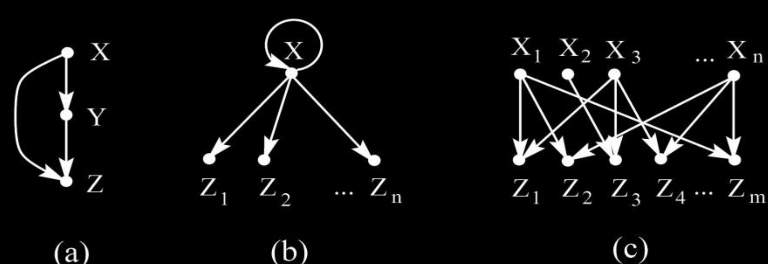
Properties of Large Networks (heuristic comparisons)

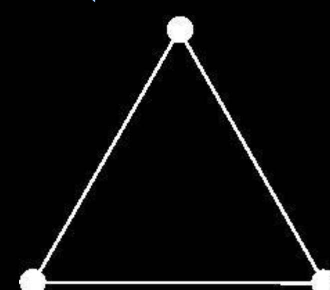
- *Global*
  - Degree distribution
  - Diameter
  - Clustering coefficient/spectrum





- *Local:*
  - network "motifs" and subgraphs (U. Alon's group, '02-'04)





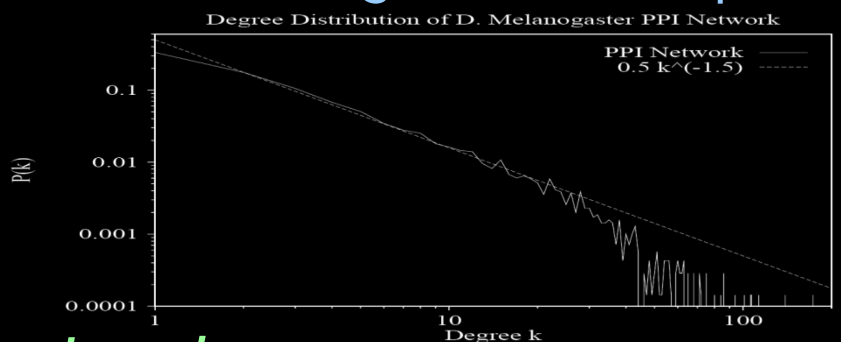
Induced vs. Partial Subgraphs

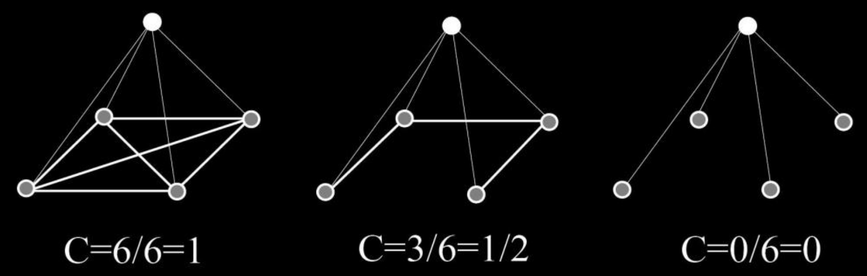
Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

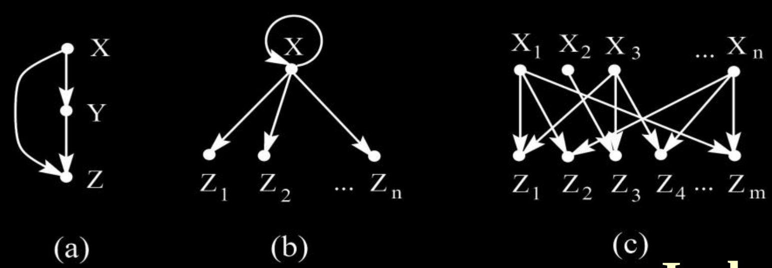
Properties of Large Networks (heuristic comparisons)

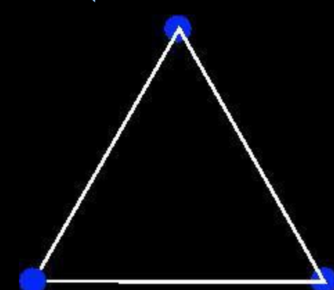
- *Global*
  - Degree distribution
  - Diameter
  - Clustering coefficient/spectrum





- *Local:*
  - network "motifs" and subgraphs (U. Alon's group, '02-'04)





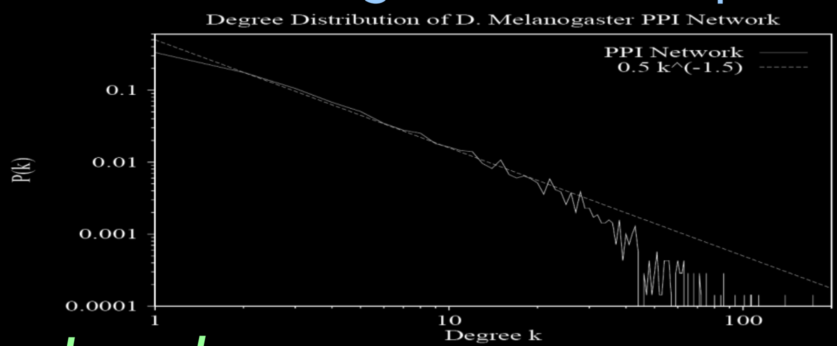
Induced vs. Partial Subgraphs

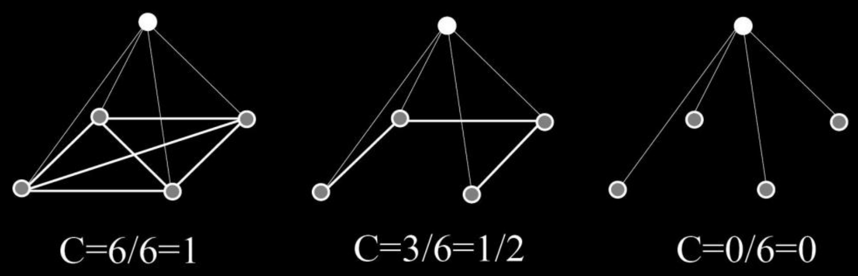
Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

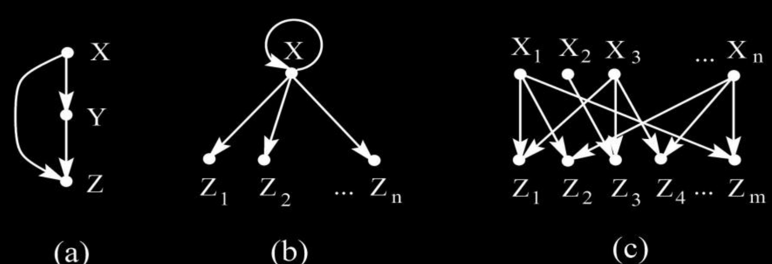
Properties of Large Networks (heuristic comparisons)

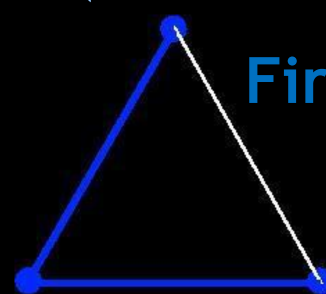
- *Global*
  - Degree distribution
  - Diameter
  - Clustering coefficient/spectrum





- *Local:*
  - network "motifs" and subgraphs (U. Alon's group, '02-'04)





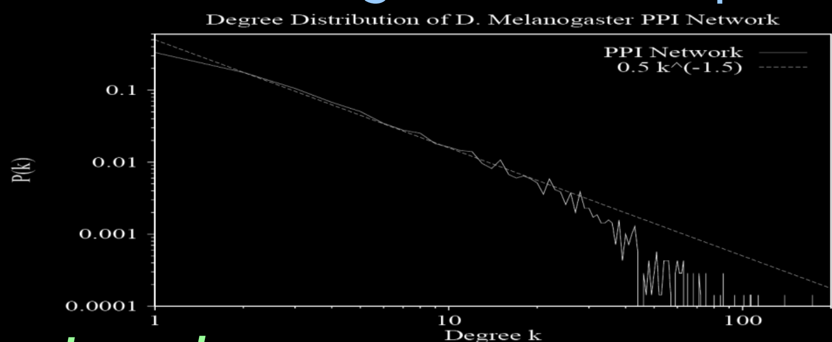
Induced vs. Partial Subgraphs

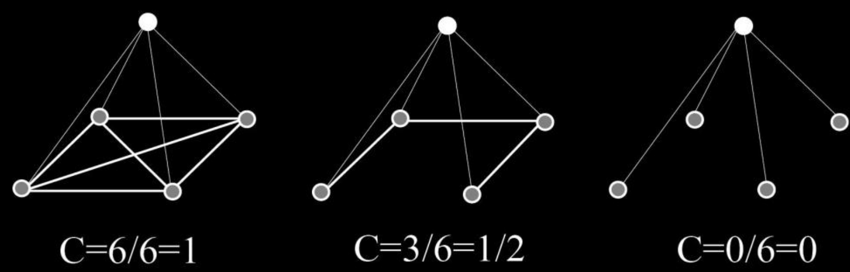
Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

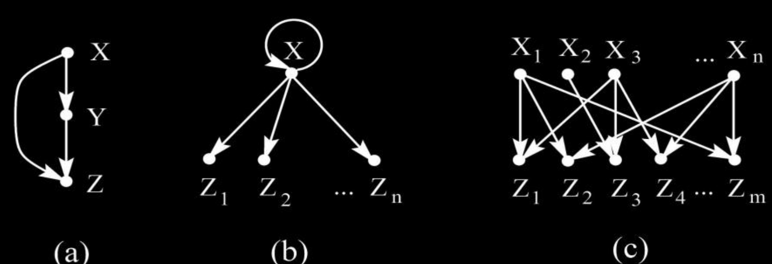
Properties of Large Networks (heuristic comparisons)

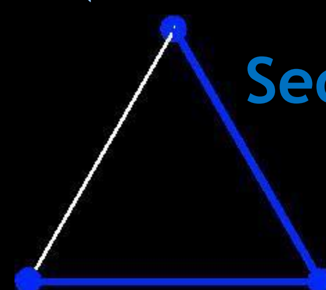
- *Global*
  - Degree distribution
  - Diameter
  - Clustering coefficient/spectrum





- *Local:*
  - network "motifs" and subgraphs (U. Alon's group, '02-'04)





Second 3-node path

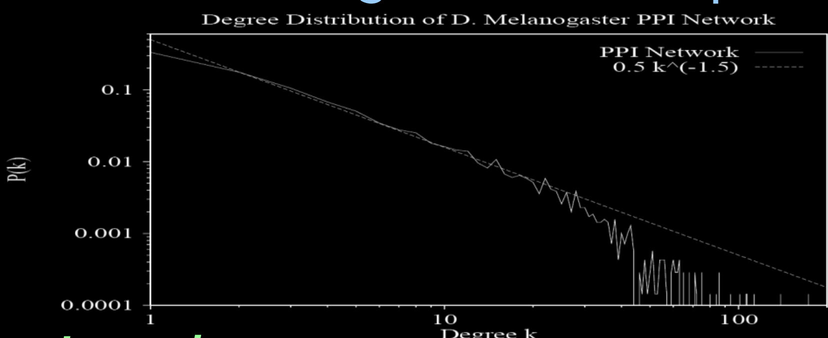
Induced vs. Partial Subgraphs

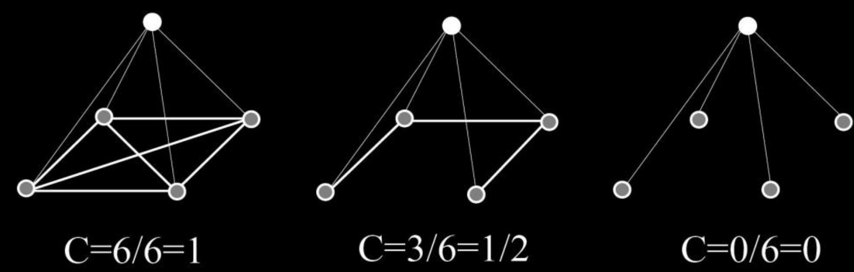
Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

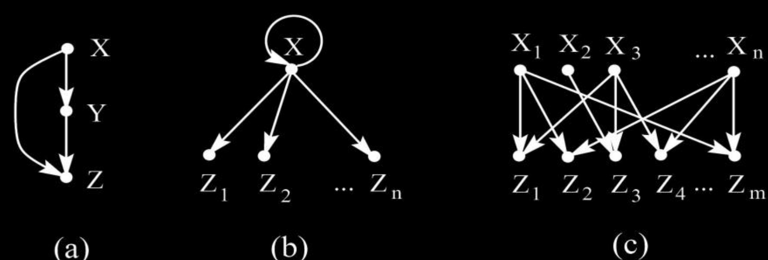
Properties of Large Networks (heuristic comparisons)

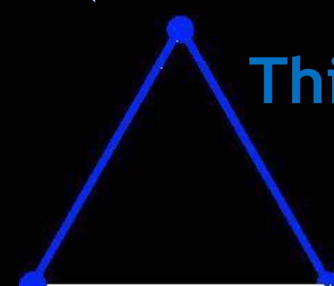
- *Global*
  - Degree distribution
  - Diameter
  - Clustering coefficient/spectrum





- *Local:*
  - network "motifs" and subgraphs (U. Alon's group, '02-'04)





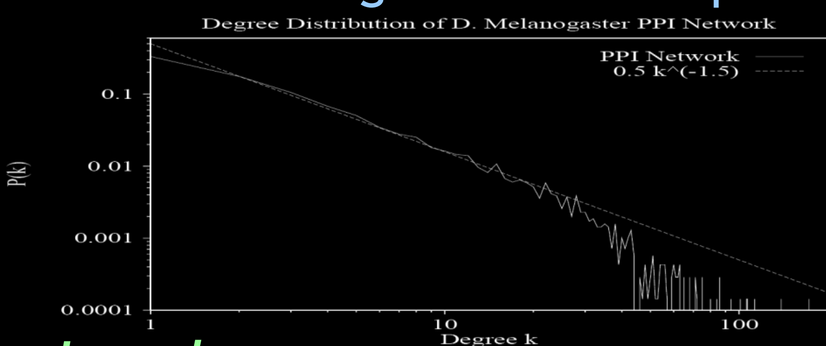
Induced vs. Partial Subgraphs

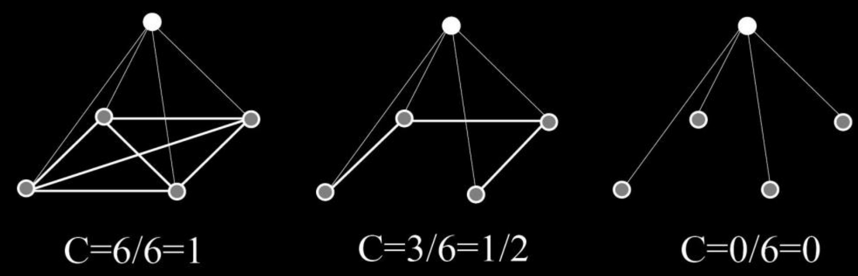
Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background

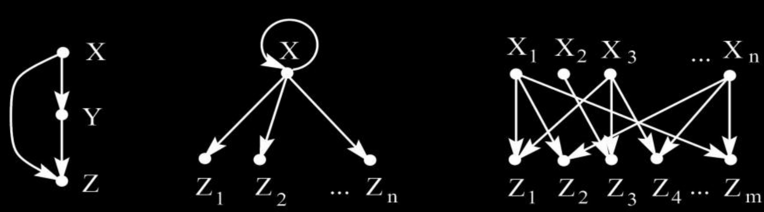
Properties of Large Networks (heuristic comparisons)

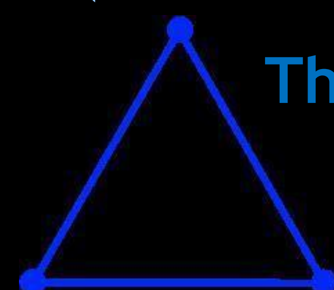
- *Global*
  - Degree distribution
  - Diameter
  - Clustering coefficient/spectrum





- *Local:*
  - network "motifs" and subgraphs (U. Alon's group, '02-'04)





The only triangle

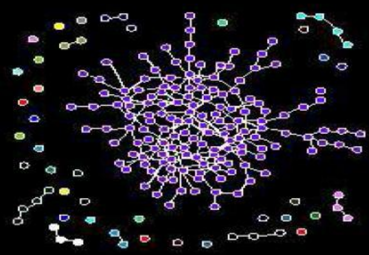
Induced vs. Partial Subgraphs

Nataša Pržulj
natasha@imperial.ac.uk

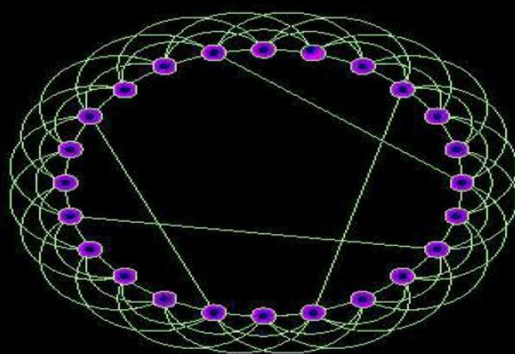
1. Introduction and Background

Examples of different model networks:

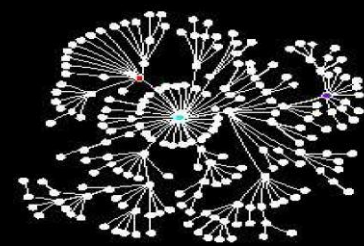
Erdös-Rényi (ER)



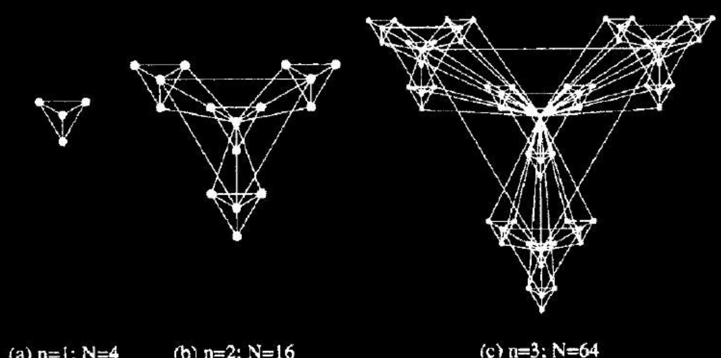
Small-World



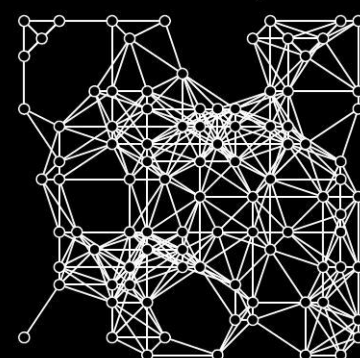
Scale-Free (SF)



Hierarchical



Geometric (GEO)

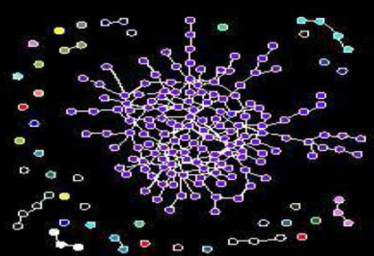


Nataša Pržulj
natasha@imperial.ac.uk

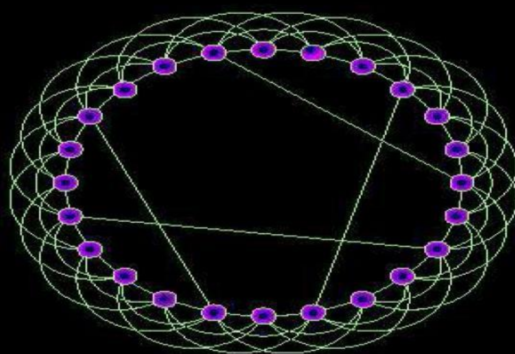
1. Introduction and Background

Examples of different model networks:

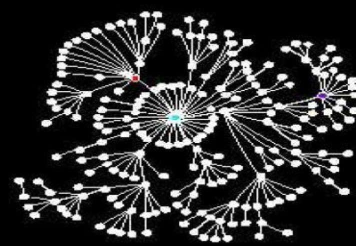
Erdös-Rényi (ER)



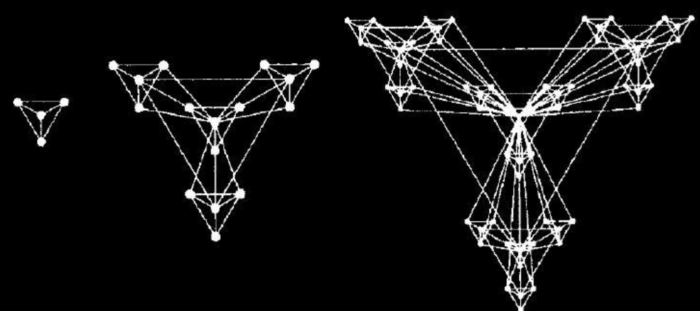
Small-World



Scale-Free (SF)

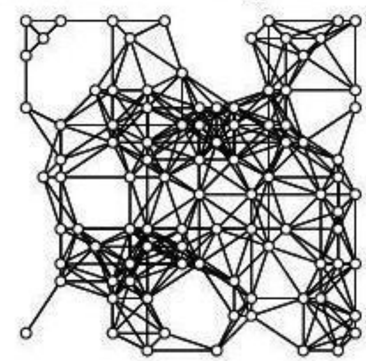


Hierarchical



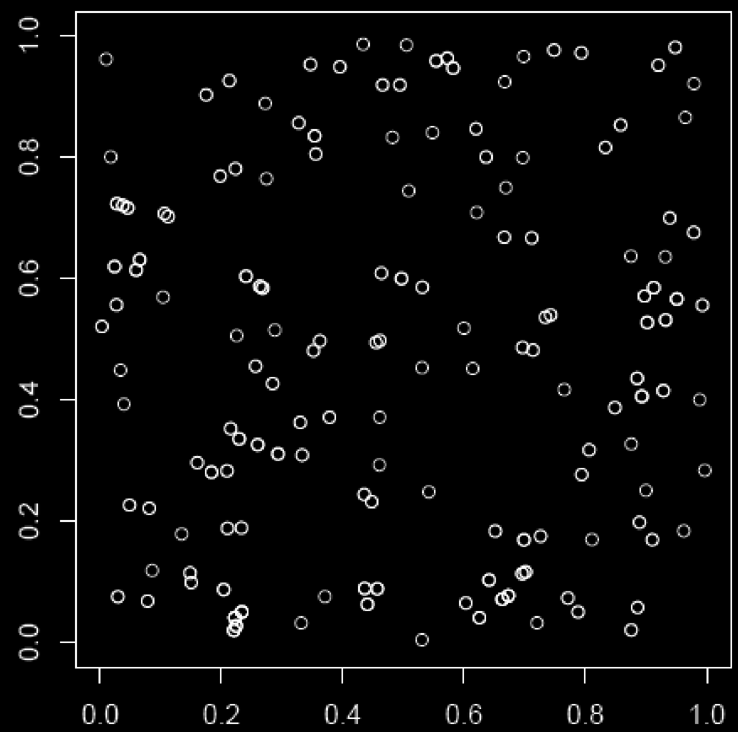
(a) n=1: N=4 (b) n=2: N=16 (c) n=3: N=64

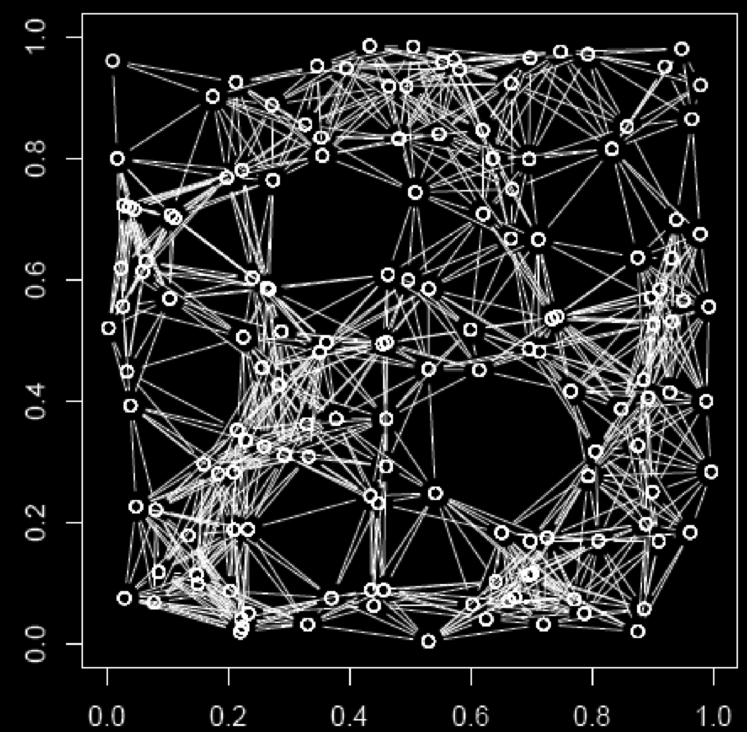
Geometric (GEO)



Nataša Pržulj
natasha@imperial.ac.uk

1. Introduction and Background





Geometric random graphs, generated using the Euclidean distance.

Nataša Pržulj
natasha@imperial.ac.uk

Overview

1. Introduction and Background:
   - PPI and other networks
   - Network analysis and modeling challenges
2. New Network Analysis and Modeling:
   (A) Analysis:
   - ➢ Structure *vs.* biological function – Graphlet Signatures
   - ➢ Network alignment

   (B) Modeling – Geometric Graphs:
   - ➢ Graphlet Degree Distributions
   - ➢ Network embedding
   - ➢ Trained Geometric Model
   - ➢ Application: De-noising PPI networks

   (C) Why PPI networks might be geometric?
3. Conclusions

Nataša Pržulj
natasha@imperial.ac.uk

Overview

1. Introduction and Background:
   - PPI and other networks
   - Network analysis and modeling challenges
2. New Network Analysis and Modeling:
   (A) Analysis:
   - Struc
   - Network alignment

   (B) Modeling – Geometric Graphs:
   - Graphlet Degree Distributions
   - Network embedding
   - Trained Geometric Model
   - Application: De-noising PPI networks

   (C) Why PPI networks might be geometric?
3. Conclusions

All robust to noise

Nataša Pržulj
natasha@imperial.ac.uk

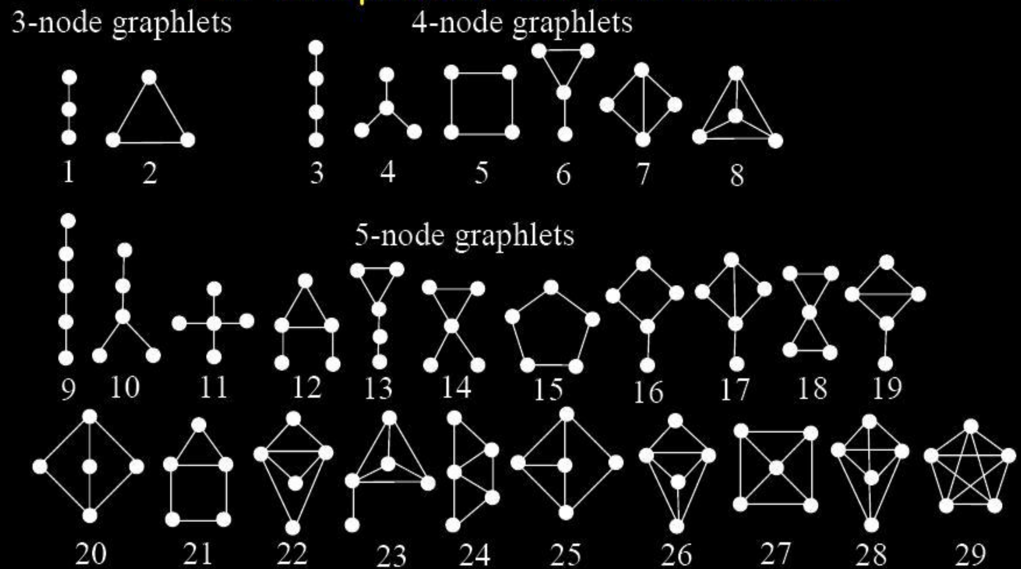
2. Network Analysis and Modeling

New measures of network local structure

Definition 3 Graphlets *are small connected non-isomorphic subgraphs of a graph G induced on $n \geq 3$ nodes of G.*

For $n = 3, 4, 5, \ldots, 10$, there are $2, 6, 21, \ldots, 11716571$ graphlets!

All Graphlets on 3-5 nodes:



N. Przulj, D. G. Corneil, and I. Jurisica, "Modeling Interactome: Scale Free or Geometric?," *Bioinformatics*, vol. 20, num. 18, pg. 3508-3515, 2004.

Nataša Pržulj
natasha@imperial.ac.uk

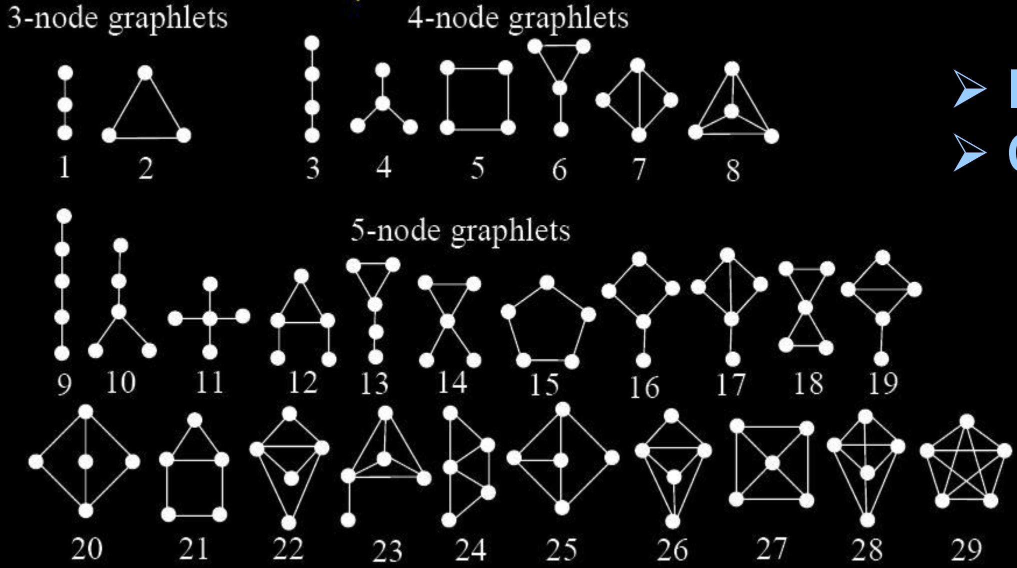
2. Network Analysis and Modeling

New measures of network local structure

Definition 3 Graphlets *are small connected non-isomorphic sub-graphs of a graph G induced on $n \geq 3$ nodes of G.*

For $n = 3, 4, 5, \ldots, 10$, there are $2, 6, 21, \ldots, 11716571$ graphlets!

All Graphlets on 3-5 nodes:



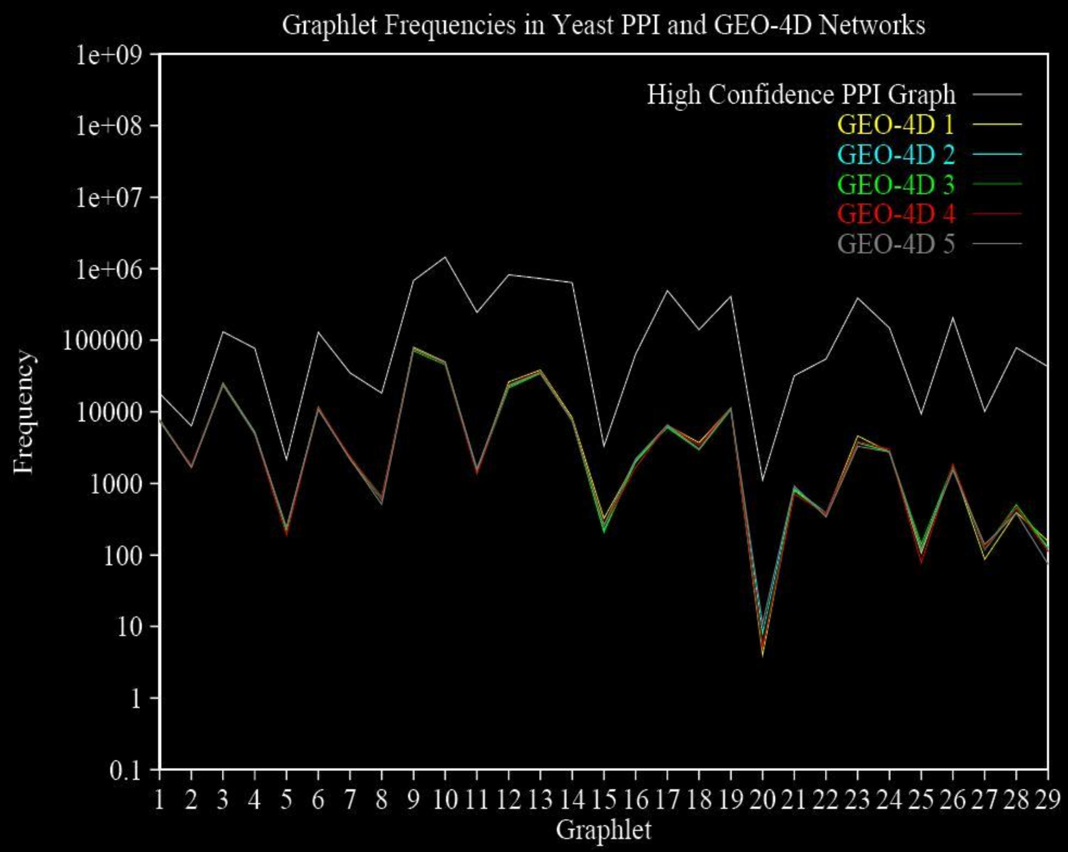
- Induced
- Of any frequency

N. Przulj, D. G. Corneil, and I. Jurisica, "Modeling Interactome: Scale Free or Geometric?," *Bioinformatics*, vol. 20, num. 18, pg. 3508-3515, 2004.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Graphlet Frequencies: *S.cerevisiae* High-Confidence PPI Network (von Mering *et al.*, *Nature 417*)

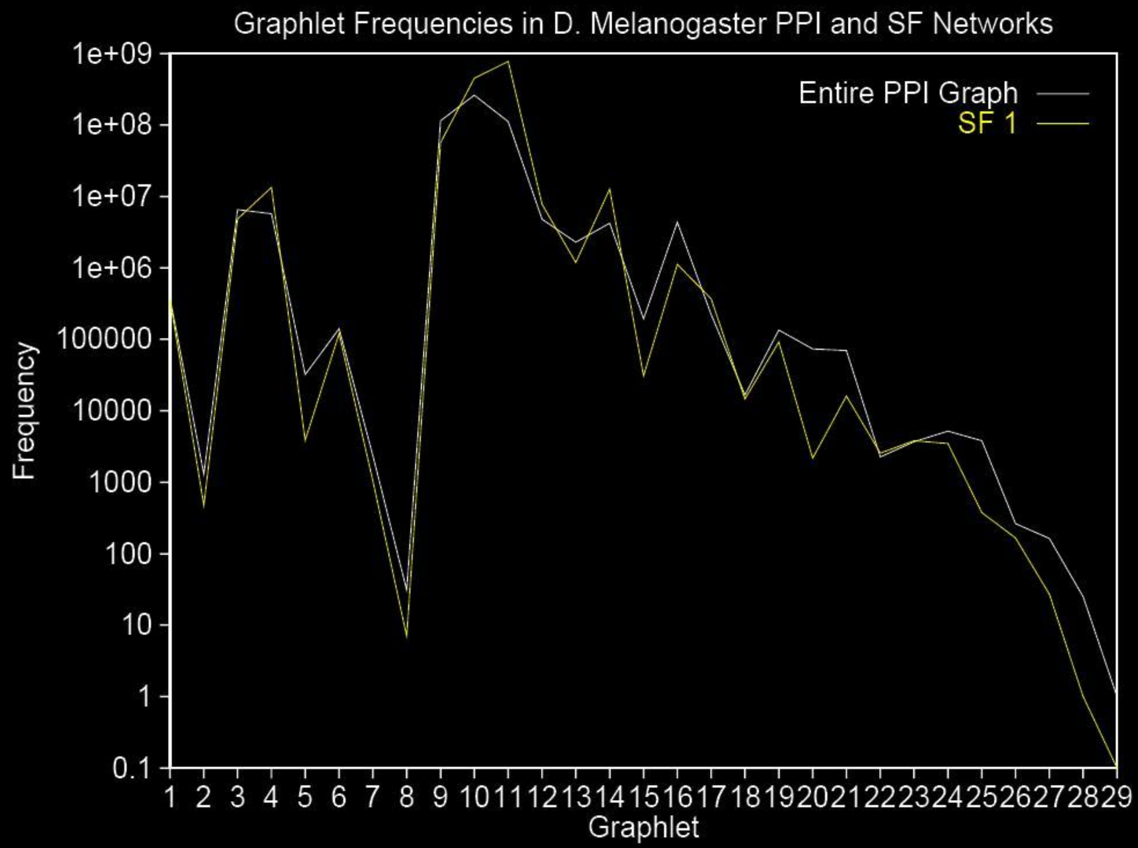


N. Przulj, D. G. Corneil, and I. Jurisica, "Modeling Interactome: Scale Free or Geometric?," *Bioinformatics*, vol. 20, num. 18, pg. 3508-3515, 2004.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Graphlet Frequencies: *D. melanogaster* Noisy PPI Network
(Giot *et al.*, *Science 302*) (77% of edges are of low confidence)

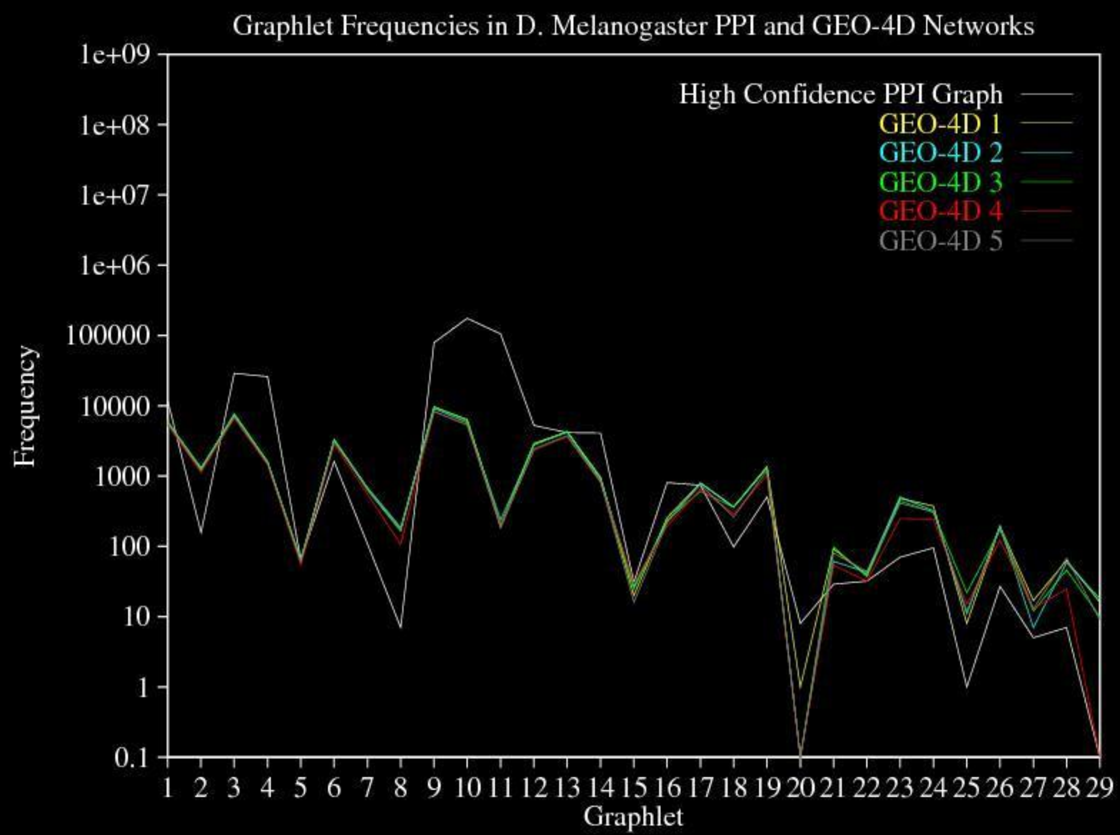


N. Przulj, D. G. Corneil, and I. Jurisica, "Modeling Interactome: Scale Free or Geometric?," *Bioinformatics*, vol. 20, num. 18, pg. 3508-3515, 2004.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Graphlet Frequencies: *D. melanogaster* Higher-Confidence PPI Network
(Giot *et al.*, *Science 302*)



N. Przulj, D. G. Corneil, and I. Jurisica, "Modeling Interactome: Scale Free or Geometric?," *Bioinformatics*, vol. 20, num. 18, pg. 3508-3515, 2004.

Nataša Pržulj
natasha@imperial.ac.uk

Imperial College London

Overview

1. Introduction and Background:
   - PPI and other networks
   - Network analysis and modeling challenges
2. New Network Analysis and Modeling:

   (A) Analysis:
   - ➢ Structure *vs.* biological function
   - ➢ Network alignment

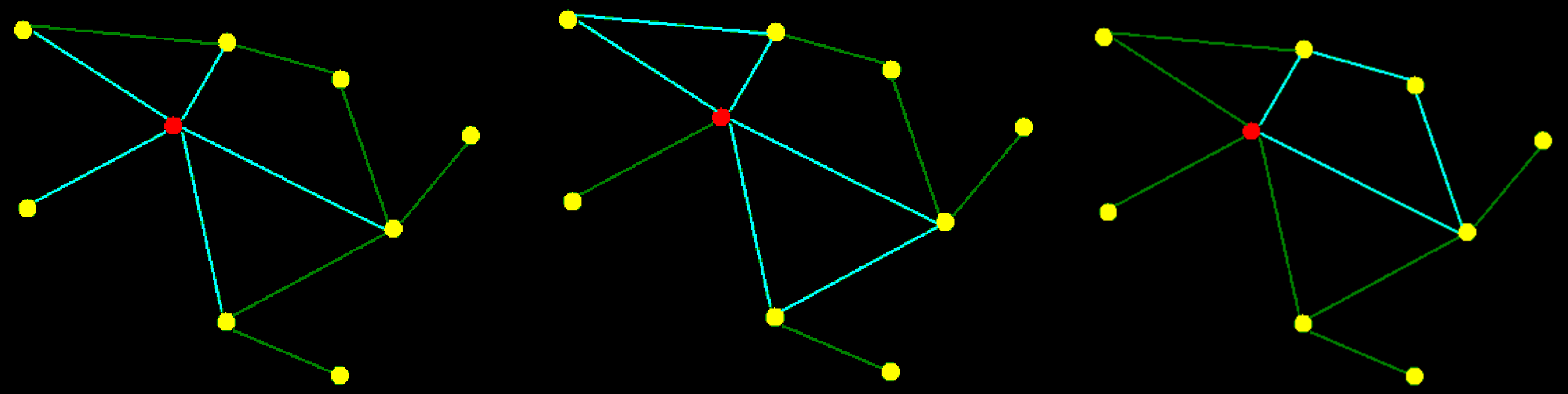
   (B) Modeling – Geometric Graphs:
   - ➢ Graphlet Degree Distributions
   - ➢ Network embedding
   - ➢ Trained Geometric Model
   - ➢ Application: De-noising PPI networks

   (C) Why PPI networks might be geometric?
3. Conclusions

Nataša Pržulj
natasha@imperial.ac.uk

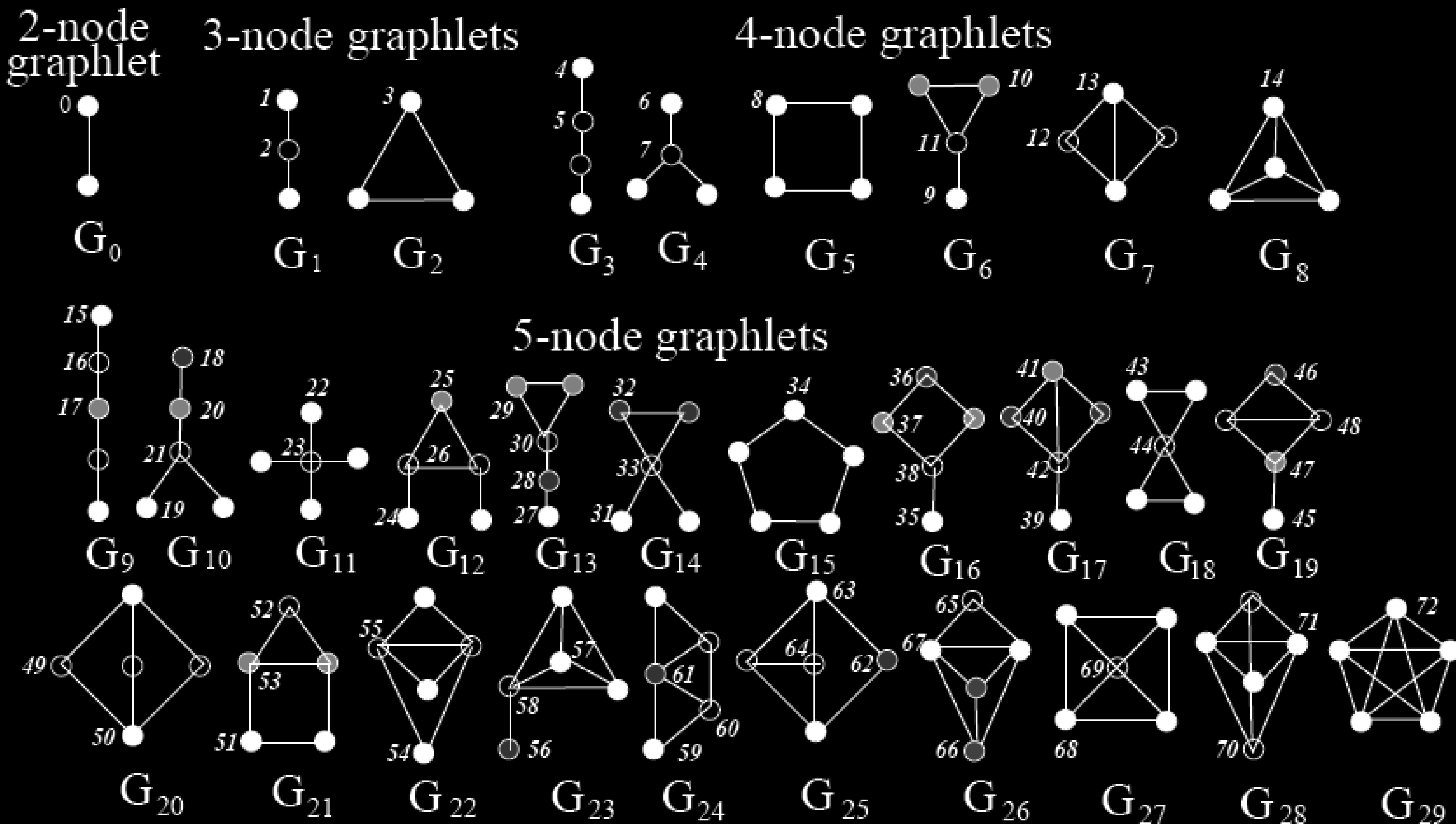
2. Network Analysis and Modeling

Generalize node **degree**



Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling



N. Przulj, "Biological Network Comparison Using Graphlet Degree Distribution," *ECCB*, *Bioinformatics*, vol. 23, pg. e177-e183, 2007.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Definitions:

An *isomorphism* f from graph G to graph H is a bijection: $f : V(G) \rightarrow V(H)$ such that xy is an edge of G iff $f(x)f(y)$ is an edge of H.

An *automorphism* is an isomorphism from a graph to itself.

The automorphisms of a graph G form a *group*, called the *automorphism group of* G, and commonly denoted by $Aut(G)$.

For a node x of graph G, the *automorphism orbit of* x is $Orb(x) = \{y \in V(G) | y = f(x) \text{ for some } f \in Aut(G)\}$, where $V(G)$ is the set of nodes of graph G.

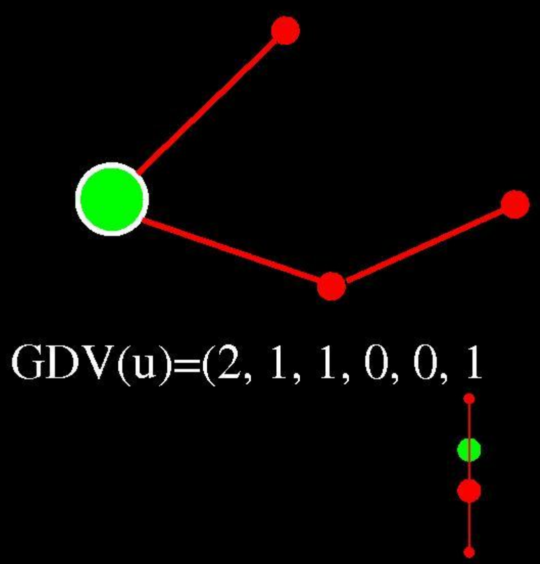
N. Przulj, "Biological Network Comparison Using Graphlet Degree Distribution," *ECCB*, *Bioinformatics*, vol. 23, pg. e177-e183, 2007.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network structure vs. biological function & disease

Graphlet Degree (GD) vectors, or "node signatures"



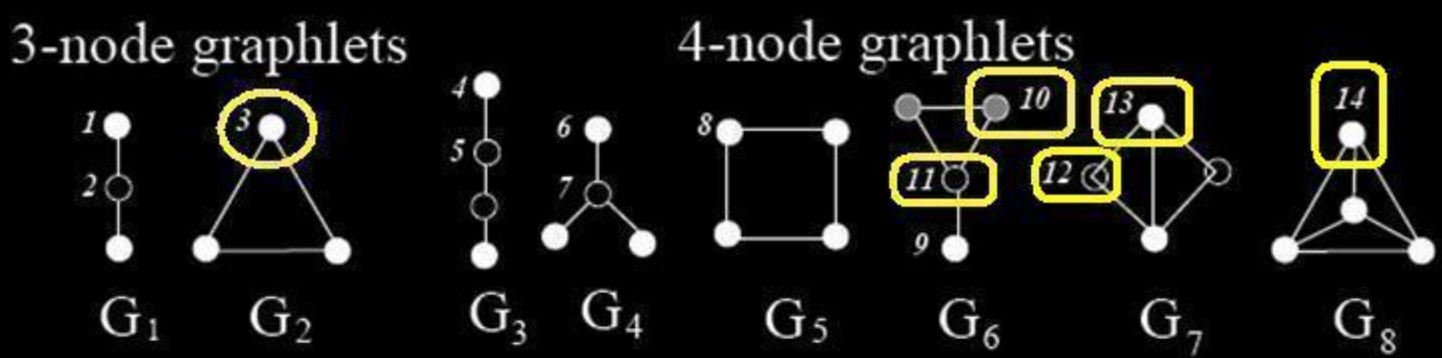
T. Milenkovic and **N. Przulj**, "Uncovering Biological Network Function via Graphlet Degree Signatures", *Cancer Informatics*, vol. 4, pg. 257-273, 2008.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network structure vs. biological function & disease

Similarity measure between nodes' Graphlet Degree vectors



T. Milenkovic and **N. Przulj**, "Uncovering Biological Network Function via Graphlet Degree Signatures", *Cancer Informatics*, vol. 4, pg. 257-273, 2008.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Signature Similarity Measure

- o_i is number of orbits that affect orbit $i \in \{0, \ldots, 72\}$
- $w_i = 1 - \frac{log(o_i)}{log(73)}$
- Distance between the i^{th} orbits of nodes u and v is

$$D_i(u,v) = w_i \times \frac{|log(u_i+1) - log(v_i+1)|}{log(max\{u_i, v_i\}+2)}$$

- The total distance between nodes u and v is

$$D(u,v) = \frac{\sum_{i=0}^{72} D_i}{\sum_{i=0}^{72} w_i}$$

- The *signature similarity* between nodes u and v is

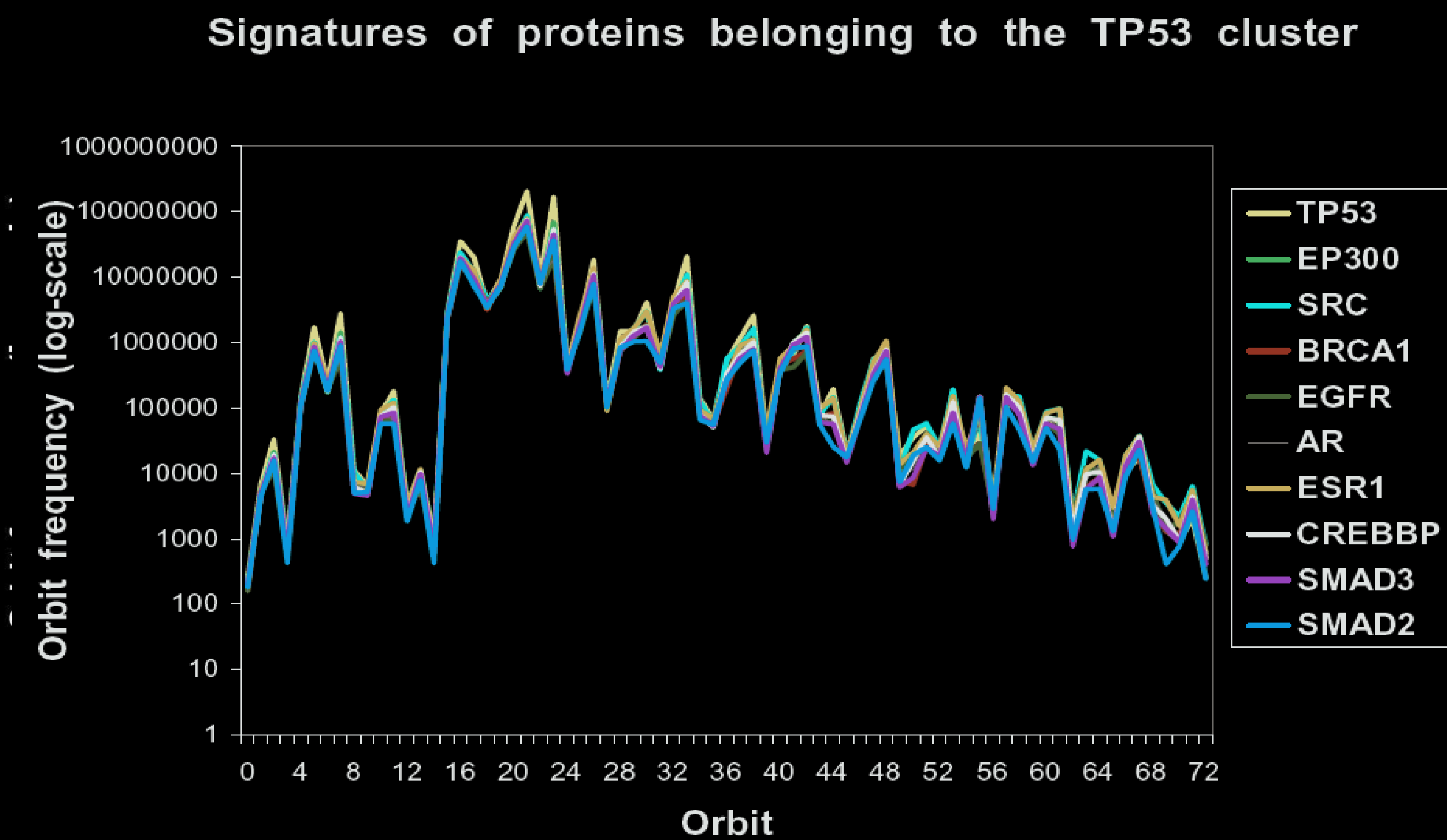
$$S(u,v) = 1 - D(u,v)$$

T. Milenkovic and **N. Przulj**, "Uncovering Biological Network Function via Graphlet Degree Signatures", *Cancer Informatics*, vol. 4, pg. 257-273, 2008.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network structure vs. biological function & disease

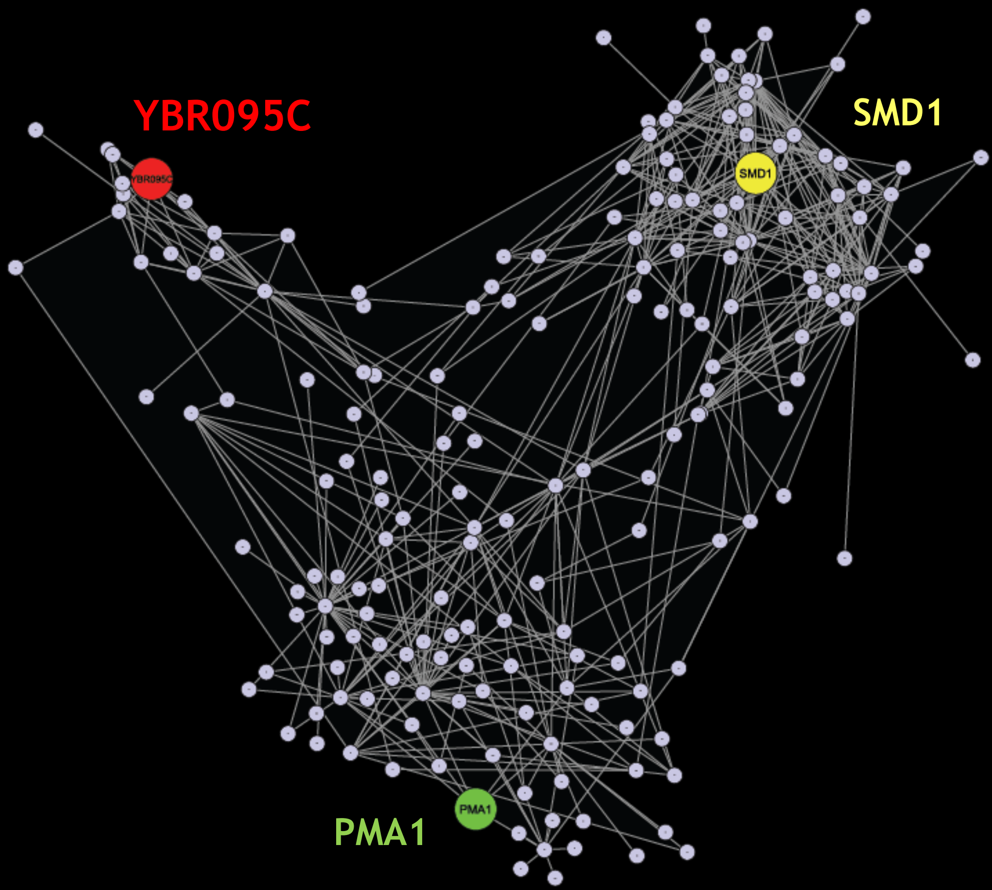


T. Milenkovic and **N. Przulj**, "Uncovering Biological Network Function via Graphlet Degree Signatures", *Cancer Informatics*, vol. 4, pg. 257-273, 2008.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network structure vs. biological function & disease



40%

T. Milenkovic and **N. Przulj**, "Uncovering Biological Network Function via Graphlet Degree Signatures", *Cancer Informatics*, vol. 4, pg. 257-273, 2008.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network structure vs. biological function & disease

T. Milenkovic and **N. Przulj**, "Uncovering Biological Network Function via Graphlet Degree Signatures", *Cancer Informatics*, vol. 4, pg. 257-273, 2008.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network structure vs. biological function & disease

- For PPI networks of yeast and human:
  - ➢ Find GD signatures for all nodes in a PPI network
  - ➢ Cluster based on signature similarities
- Obtained clusters are significantly enriched in:
  - ➢ Biological function
  - ➢ Protein complexes
  - ➢ Sub-cellular localization
  - ➢ Tissue expression
  - ➢ Disease

$\Rightarrow$ **Predict protein function and involvement in disease**

T. Milenkovic and **N. Przulj**, "Uncovering Biological Network Function via Graphlet Degree Signatures", *Cancer Informatics*, vol. 4, pg. 257-273, 2008.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network structure vs. biological function & disease

- For PPI networks of yeast and human:
  - Find GD signatures for all nodes in a PPI network
  - Cluster based on signature similarities
- Obtained clusters are significantly enriched in:
  - Biological function
  - Protein complexes
  - Sub-cellular localization
  - Tissue expression
  - Disease

⇒ Find new members of melanoma pathways

T. Milenkovic, V. Memisevic, A. K. Ganesan, and **N. Przulj**, "Systems-level Cancer Gene Identification from Protein Interaction Network Topology Applied to Melanogenesis-related Interaction Networks", *J. Roy. Soc. Interface*, July 22, 2009.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network structure vs. biological function & disease

- **For PPI networks of yeast and human:**
  - Find GD signatures for all nodes in a PPI network
  - Cluster based on signature similarities
- **Obtained clusters are significantly enriched in:**
  - Biological function
  - Protein complexes
  - Sub-cellular localization
  - Tissue expression
  - Disease

⇒ **Find new members of melanoma pathways**
could not have been identified by existing approaches

H. Ho, T. Milenkovic, V. Memisevic, J. Aruri, **N. Przulj**, and A. K. Ganesan, "Protein Interaction Network Uncovers Melanogenesis Regulatory Network Components within Functional Genomics Datasets," *BMC Systems Biology*, 4:84, June 15, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network structure vs. biological function & disease

- **For PPI networks of yeast and human:**
  - ➢ Find GD signatures for all nodes in a PPI network
  - ➢ Cluster based on signature similarities
- **Obtained clusters are significantly enriched in:**
  - ➢ Biological function
  - ➢ Protein complexes
  - ➢ Sub-cellular localization
  - ➢ Tissue expression
  - ➢ Disease

⇒ New members of the yeast proteosome PPI network

C. Guerrero, T. Milenkovic, **N. Przulj**, P. Keiser, L. Huang, "Characterization of the proteasome interaction network using a QTAX-based tag-team strategy and protein interaction network analysis," *PNAS*, 105 (36), pg. 13333-13338 2008.

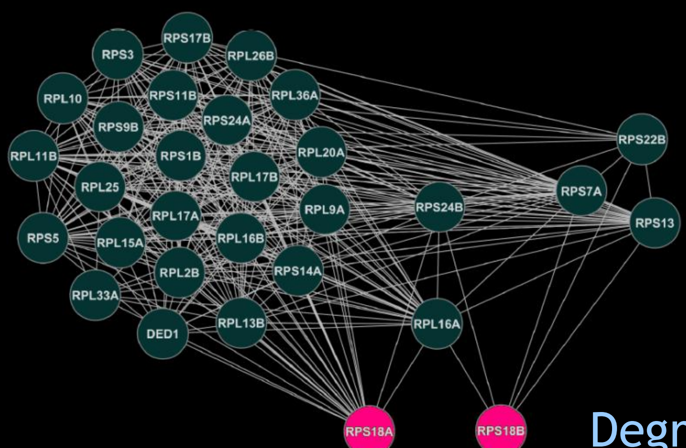
Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network structure vs. biological function & disease

Topology and Sequence: complementary sources of homology info.

- 59 of the yeast ribosomal proteins – retained two genomic copies
- Are duplicated proteins functionally redundant?
- No: have different genetic requirements for their assembly and localization, so are functionally distinct
- Also note: avg sequence identity of struct. similar prots ~8-10%
- E.g., two pairs with identical sequence:



100% sequence identity
50% signature similarity

Degrees 25 and 5

V. Memisevic, T. Milenkovic and N. Przulj, "Complementarity of network and sequence information in homologous proteins", *J. Integrative Bioinformatics*, 2010.

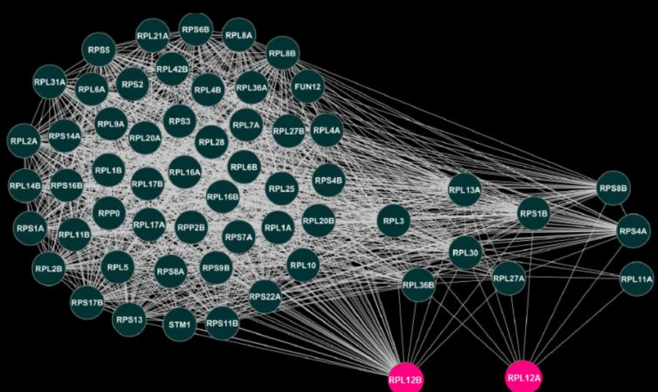
Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network structure vs. biological function & disease

Topology and Sequence: complementary sources of homology info.

- 59 of the yeast ribosomal proteins – retained two genomic copies
- Are duplicated proteins functionally redundant?
- No: have different genetic requirements for their assembly and localization, so are functionally distinct
- Also note: avg sequence identity of struct. similar prots ~8-10%
- E.g., two pairs with identical sequence:



100% sequence identity
65% signature similarity

Degrees 54 and 9

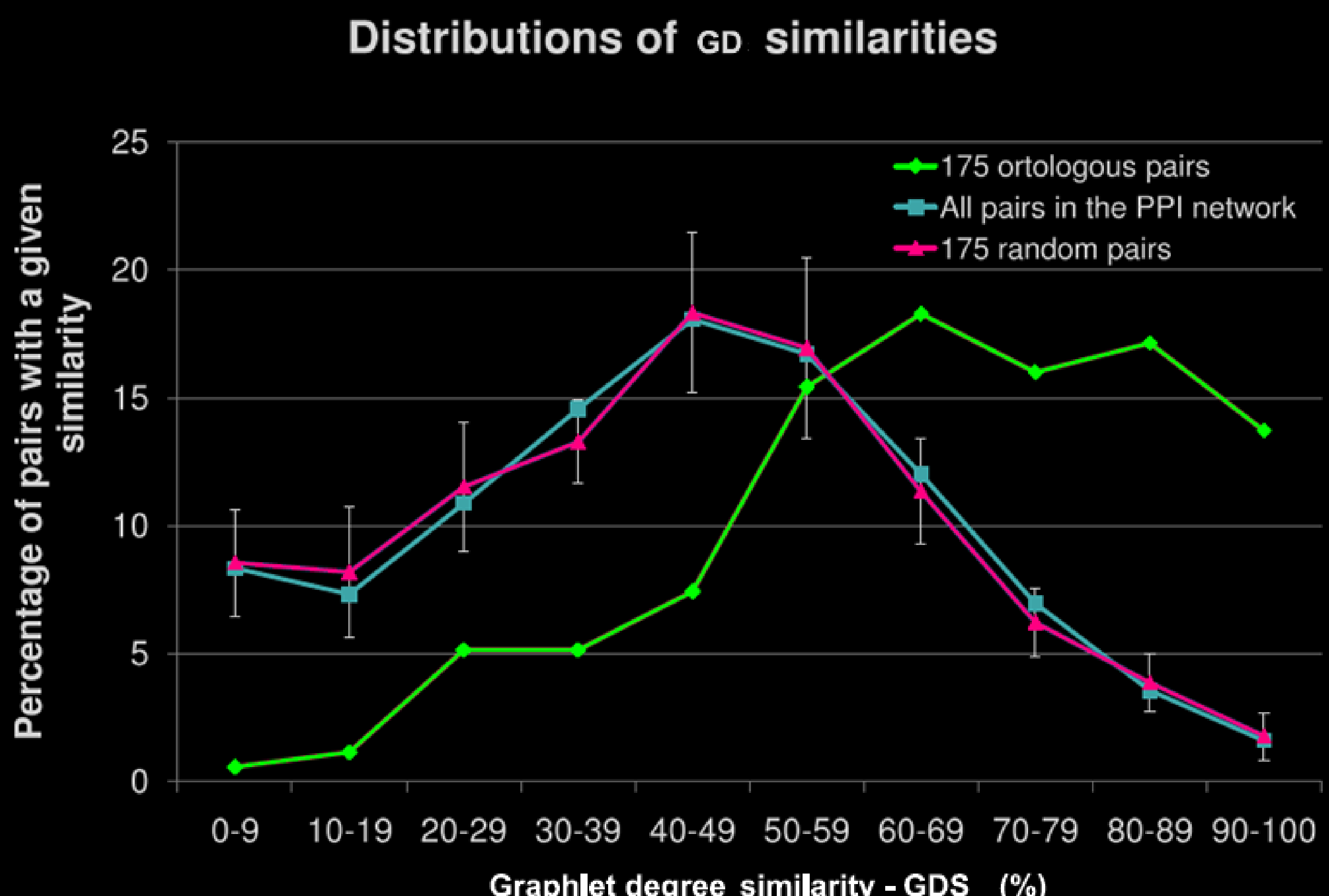
V. Memisevic, T. Milenkovic and **N. Przulj,** "Complementarity of network and sequence information in homologous proteins", *J. Integrative Bioinformatics*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Topology

- Orthologous proteins (yeast) have high GD vector similarities



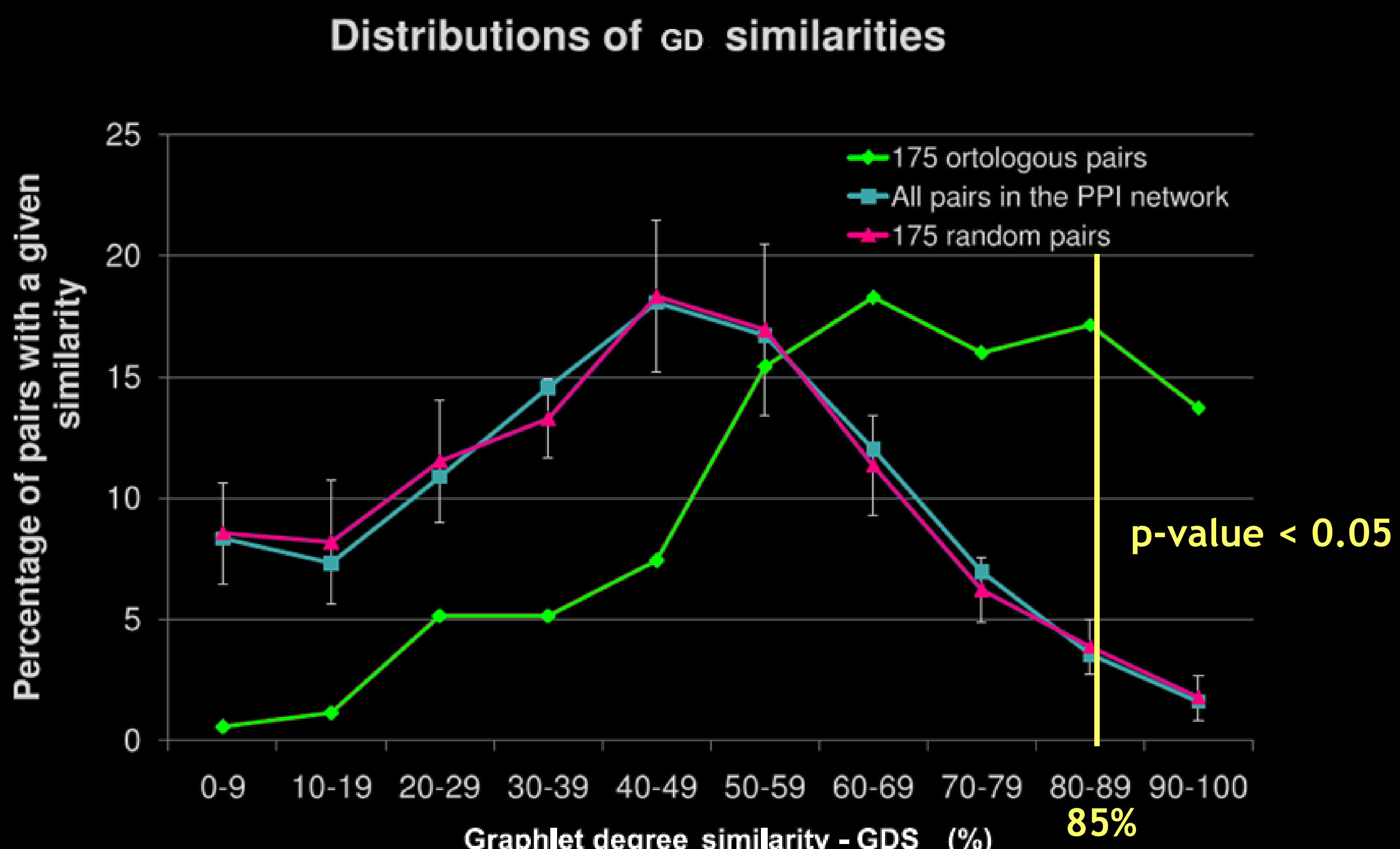
V. Memisevic, T. Milenkovic and N. Przulj, "Complementarity of network and sequence information in homologous proteins", *J. Integrative Bioinformatics*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Topology

- Orthologous proteins (yeast) have high GD vector similarities



V. Memisevic, T. Milenkovic and **N. Przulj,** "Complementarity of network and sequence information in homologous proteins", *J. Integrative Bioinformatics*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

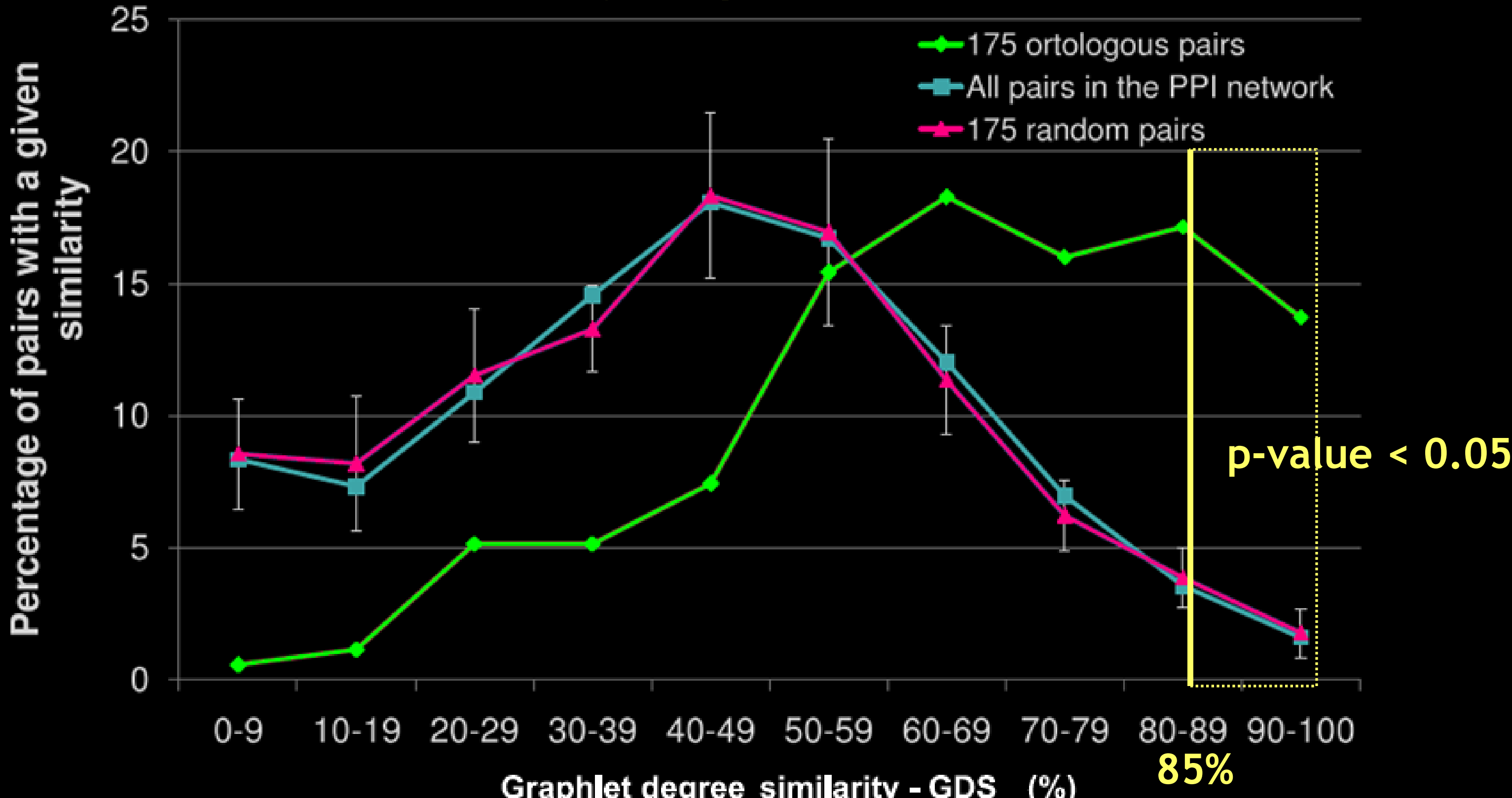
2. Network Analysis and Modeling

Network Topology

- Orthologous proteins (yeast) have high GD vector similarities

Distributions of GD similarities

> 20% of orthologous pairs have GDS > 85%



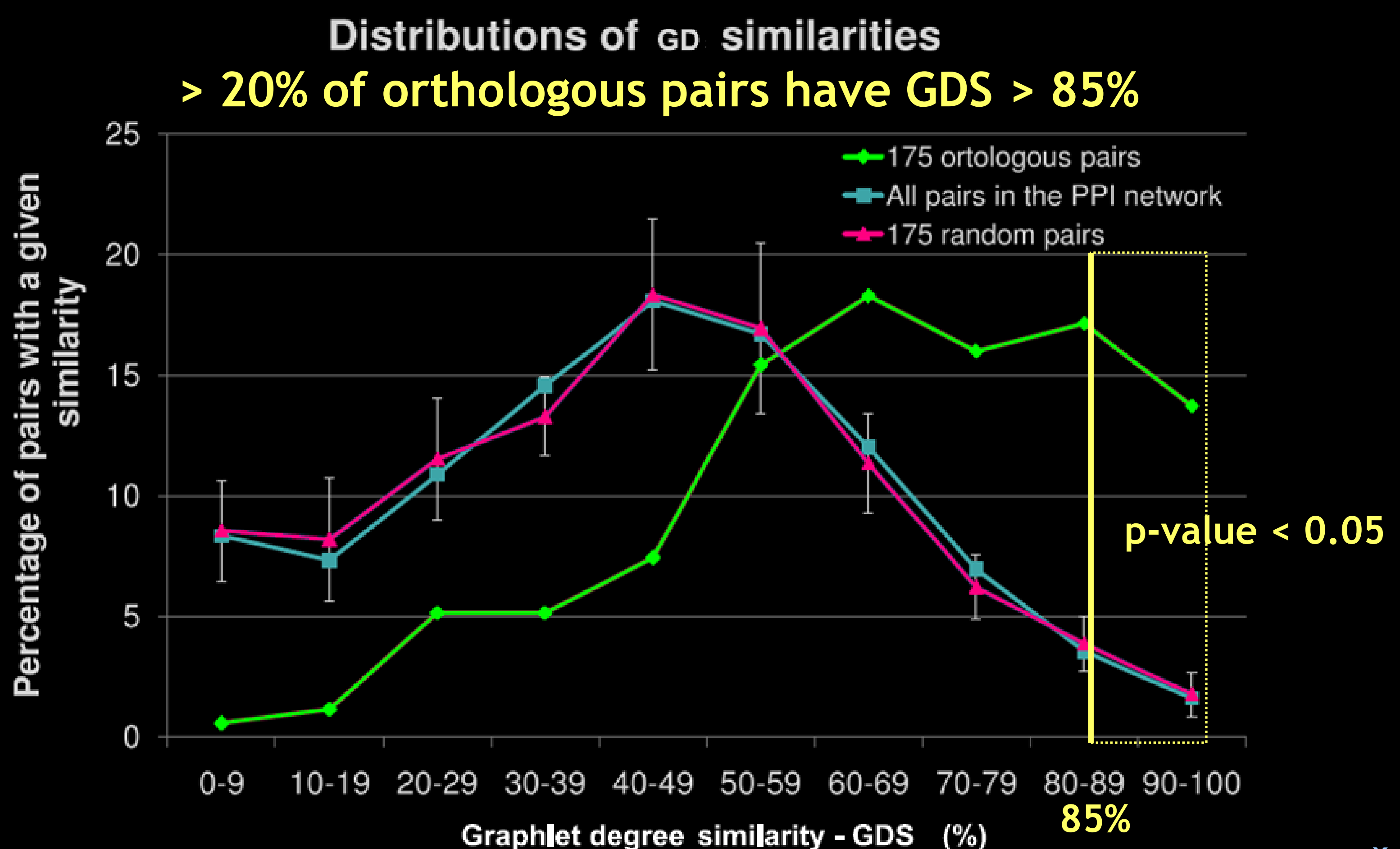
V. Memisevic, T. Milenkovic and **N. Przulj,** "Complementarity of network and sequence information in homologous proteins", *J. Integrative Bioinformatics*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

2. Netw Robust to noise

Network Topology

- Orthologous proteins (yeast) have high GD vector similarities

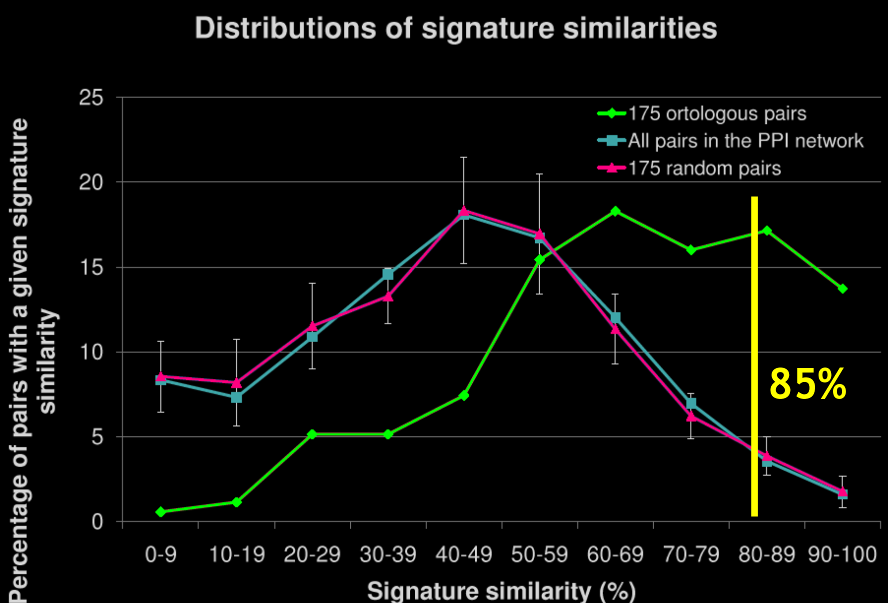


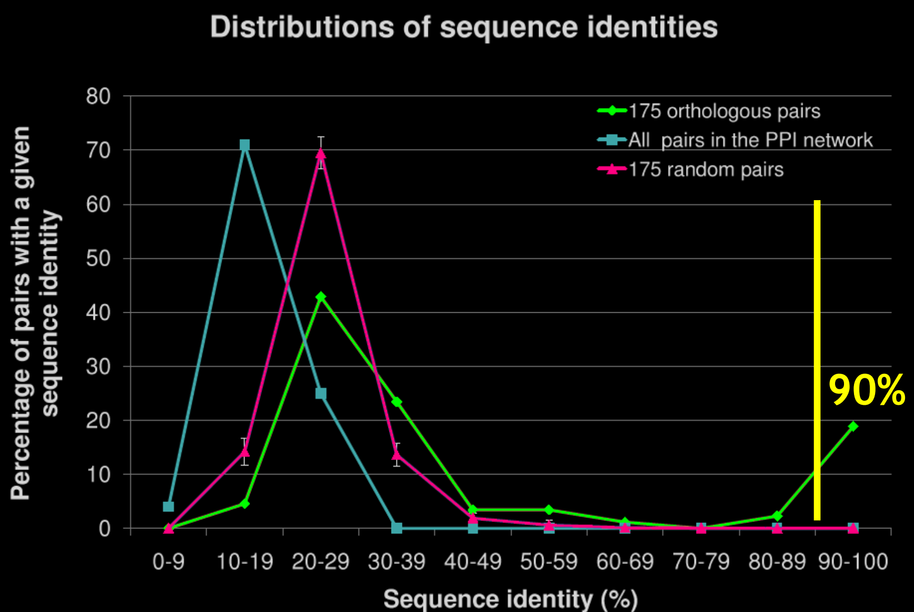
V. Memisevic, T. Milenkovic and **N. Przulj,** "Complementarity of network and sequence information in homologous proteins", *J. Integrative Bioinformatics*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Comparison of topology and sequence w.r.t. orthology:





~20% of orthologous pairs have signature similarities above 85% (35 pairs)

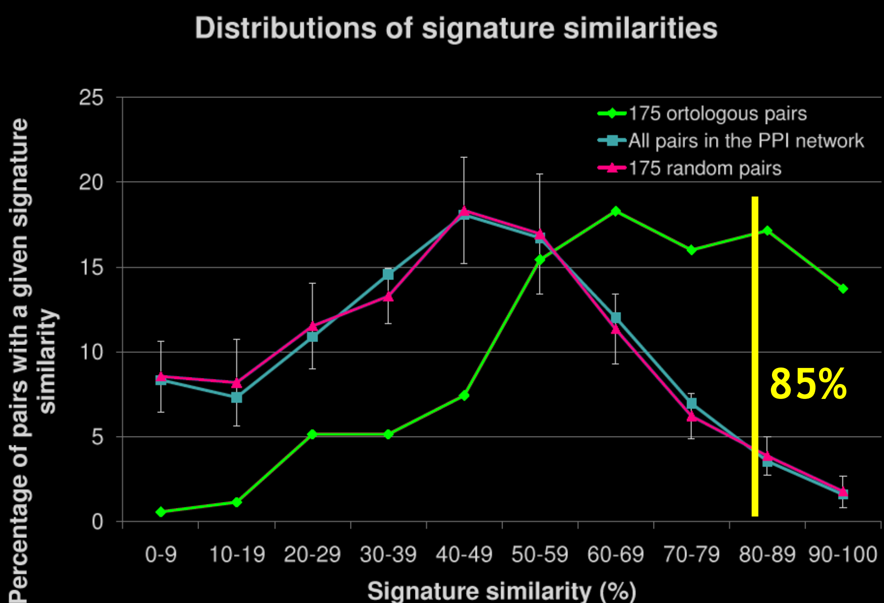
~20% orth. pairs have seq. identity > 90%

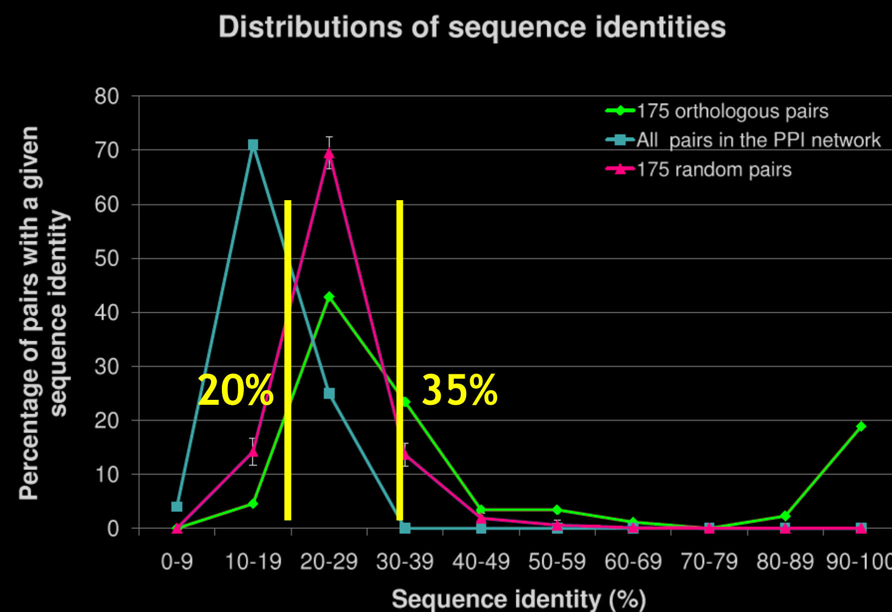
V. Memisevic, T. Milenkovic and **N. Przulj,** "Complementarity of network and sequence information in homologous proteins", *J. Integrative Bioinformatics*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Comparison of topology and sequence w.r.t. orthology:





~20% of orthologous pairs have signature similarities above 85% (35 pairs)

"Twilight zone" for homology

~70% orth. pairs have seq. identity < 35%

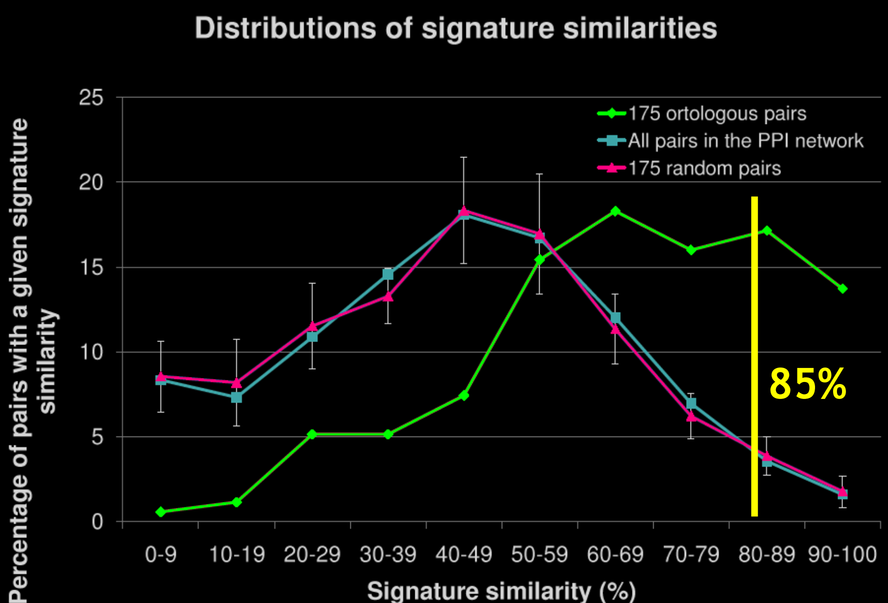
⇒ No dependence on the absolute similarity COG & KEGG, but triangles in the graph of best matches

V. Memisevic, T. Milenkovic and **N. Przulj,** "Complementarity of network and sequence information in homologous proteins", *J. Integrative Bioinformatics*, 2010.

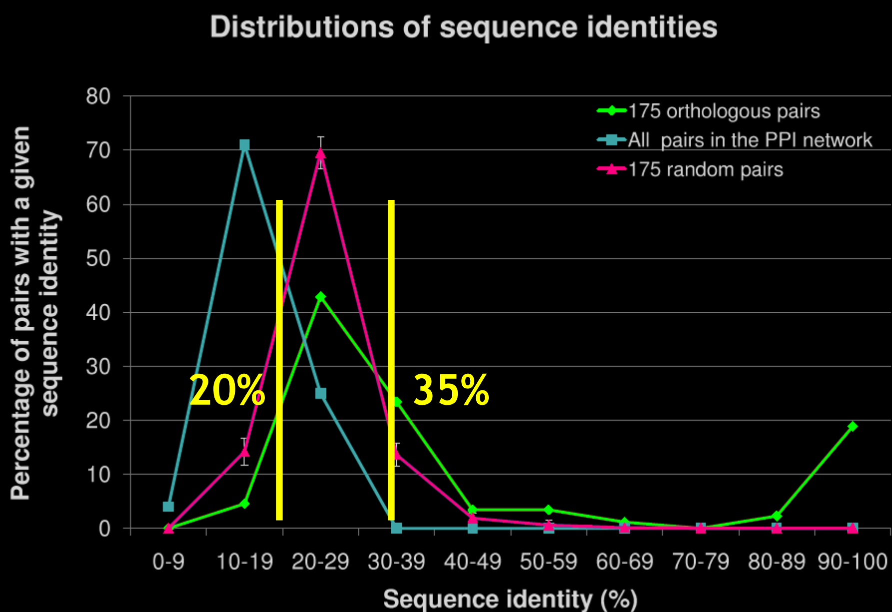
Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Comparison of topology and sequence w.r.t. orthology:



~20% of orthologous pairs have signature similarities above 85% (35 pairs)



~30% of orthologous pairs have sequence identities above 35% (53 pairs)

Overlap: 22 pairs (~60% of the smaller set)

⇒ Sequence and network topology
somewhat complementary slices of homology information

V. Memisevic, T. Milenkovic and **N. Przulj**, "Complementarity of network and sequence information in homologous proteins", *J. Integrative Bioinformatics*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

Overview

1. Introduction and Background:
   - PPI and other networks
   - Network analysis and modeling challenges
2. New Network Analysis and Modeling:

   (A) Analysis:
   - ➢ Structure *vs.* biological function
   - ➢ Network alignment

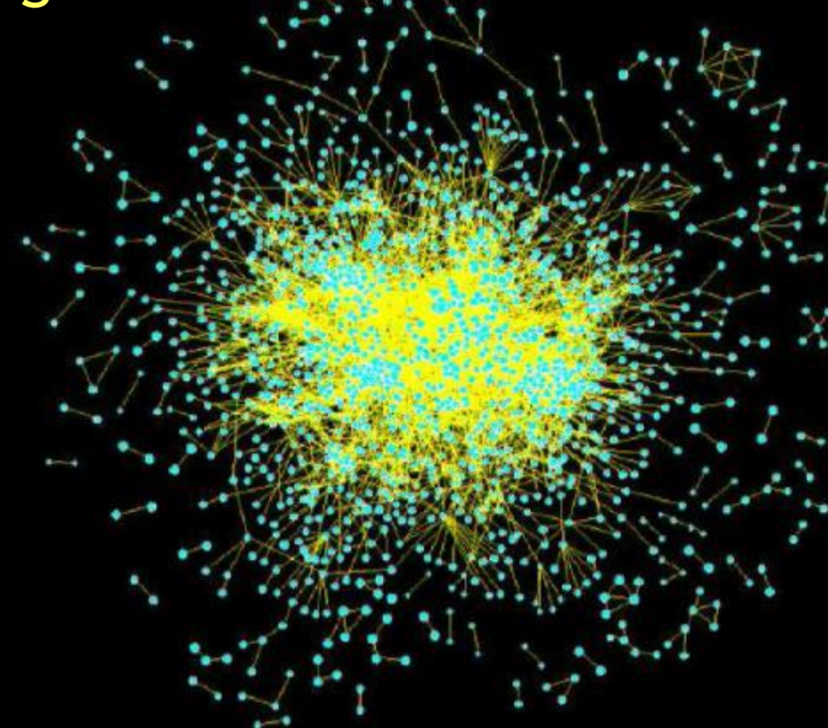
   (B) Modeling – Geometric Graphs:
   - ➢ Graphlet Degree Distributions
   - ➢ Network embedding
   - ➢ Trained Geometric Model
   - ➢ Application: De-noising PPI networks

   (C) Why PPI networks might be geometric?
3. Conclusions

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Alignment

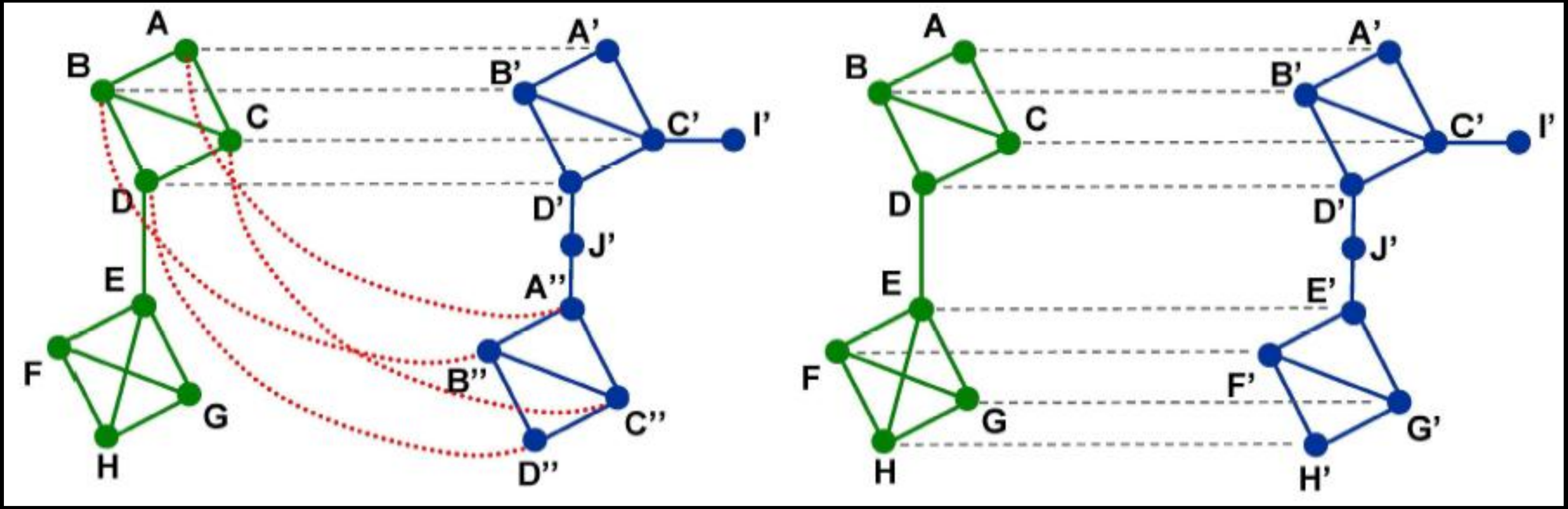


- Exact network comparisons are computationally intractable
  - ➢ Subgraph Isomorphism Problem is NP-C (Cook, 1971)
  - ➢ Rely on approximate or heuristic approaches

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Alignment



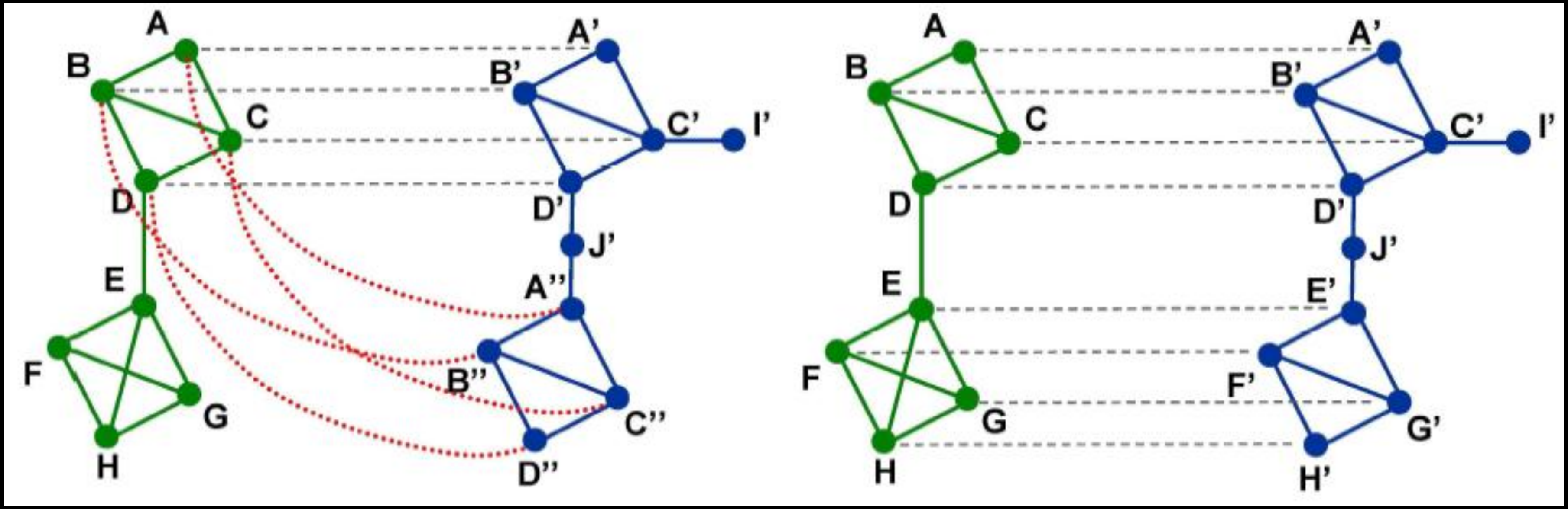
Local network alignment

Global network alignment

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Alignment



Local network alignment

Global network alignment

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Alignment

- Find **GD node signatures** across different networks
- Align "signature-similar" nodes – **"seed nodes"** in each network
- **Extend** around seed nodes
  - Greedy seed-and-extend approach
  - Align **PPI networks** of yeast and human

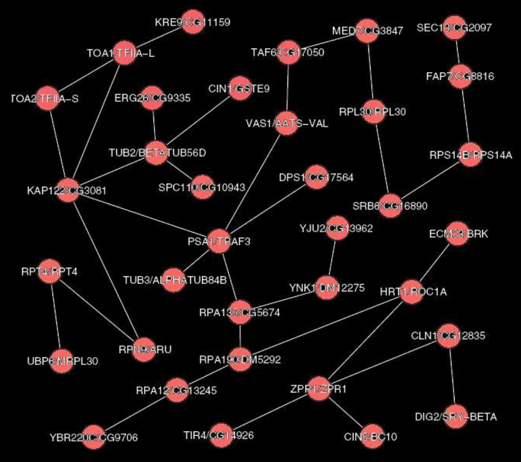
O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes and **N. Przulj**, "Topological Network Alignment Uncovers Biological Function and Phylogeny", *J. Roy Soc Interface*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Alignment

- Find GD node signatures across different networks
- Align "signature-similar" nodes – "seed nodes" in each network
- Extend around seed nodes
  - ➢ Greedy seed-and-extend approach
  - ➢ Align PPI networks of yeast and human



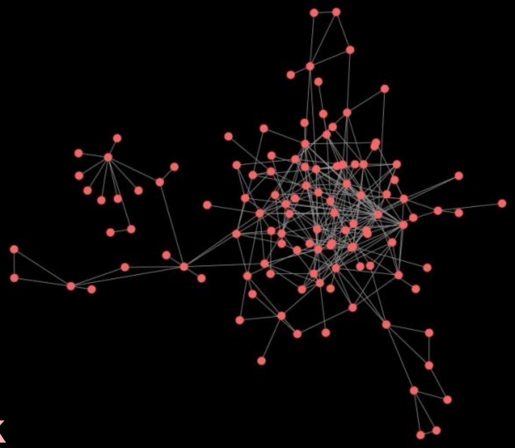
Isorank: Singh, Xu, Berger, "Pairwise Global Alignment of Protein Interaction Netowrks by Matching Neighborhood Topology," *RECOMB* 2007, LNBI 4453, pp. 1631, 2007.

O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes and **N. Przulj**, "Topological Network Alignment Uncovers Biological Function and Phylogeny", *J. Roy Soc. Interface*, 2010.

2. Network Analysis and Modeling

Network Alignment

- Find GD node signatures across different networks
- Align “signature-similar” nodes – “seed nodes” in each network
- Extend around seed nodes
  - ➢ Greedy seed-and-extend approach
  - ➢ Align PPI networks of yeast and human



Isorank

Isorank: Singh, Xu, Berger, “Pairwise Global Alignment of Protein Interaction Netowrks by Matching Neighborhood Topology,” *RECOMB* 2007, LNBI 4453, pp. 1631, 2007.

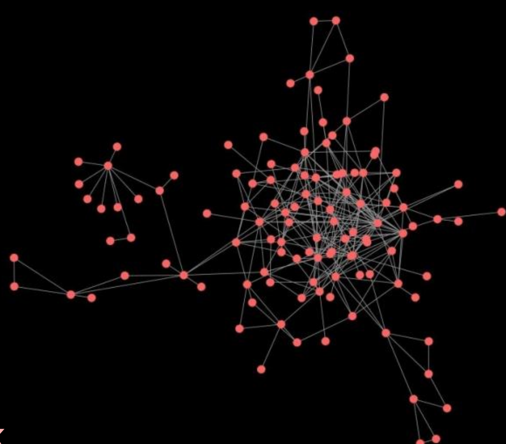
O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes and **N. Przulj**, “Topological Network Alignment Uncovers Biological Function and Phylogeny”, *J. Roy Soc. Interface*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

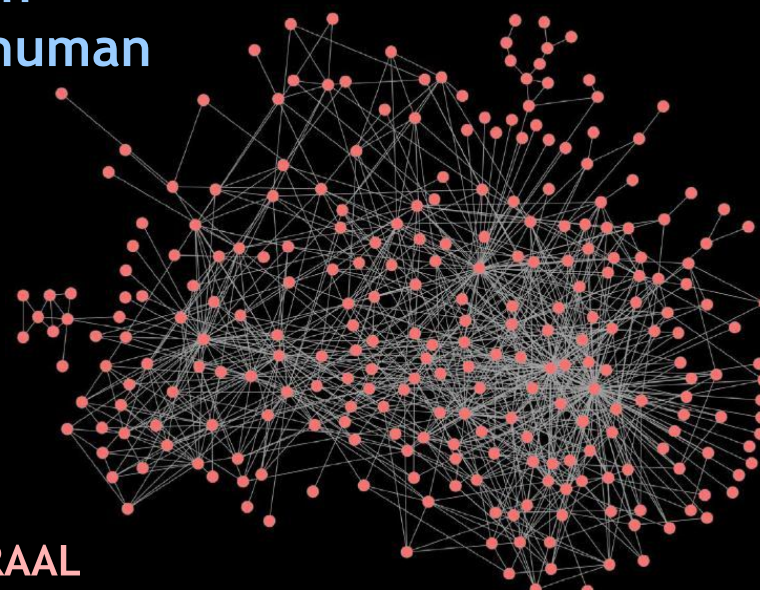
2. Network Analysis and Modeling

Network Alignment

- Find GD node signatures across different networks
- Align "signature-similar" nodes – "seed nodes" in each network
- Extend around seed nodes
  - ➢ Greedy seed-and-extend approach
  - ➢ Align PPI networks of yeast and human



Isorank



GRAAL
(GRAph ALigner)

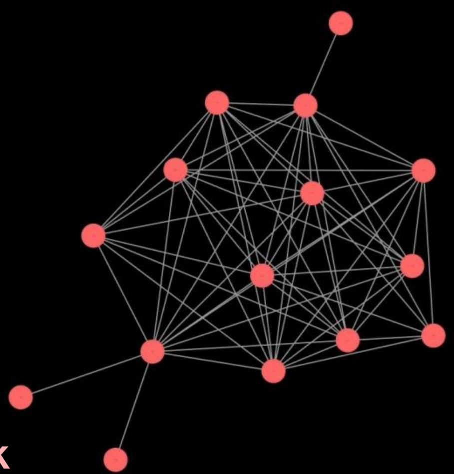
O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes and **N. Przulj**, "Topological Network Alignment Uncovers Biological Function and Phylogeny", *J. Roy Soc. Interface*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Alignment

- Find GD node signatures across different networks
- Align "signature-similar" nodes – "seed nodes" in each network
- Extend around seed nodes
  - Greedy seed-and-extend approach
  - Align PPI networks of yeast and human



Isorank

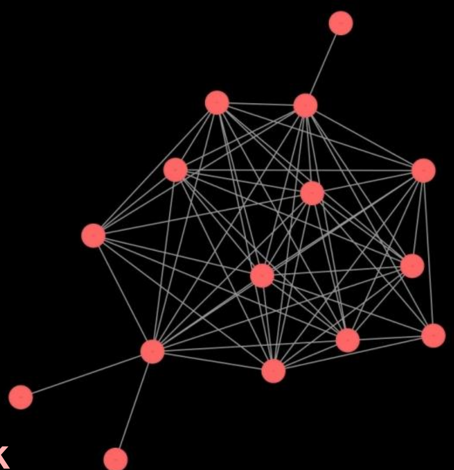
O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes and **N. Przulj**, "Topological Network Alignment Uncovers Biological Function and Phylogeny", *J. Roy Soc. Interface*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

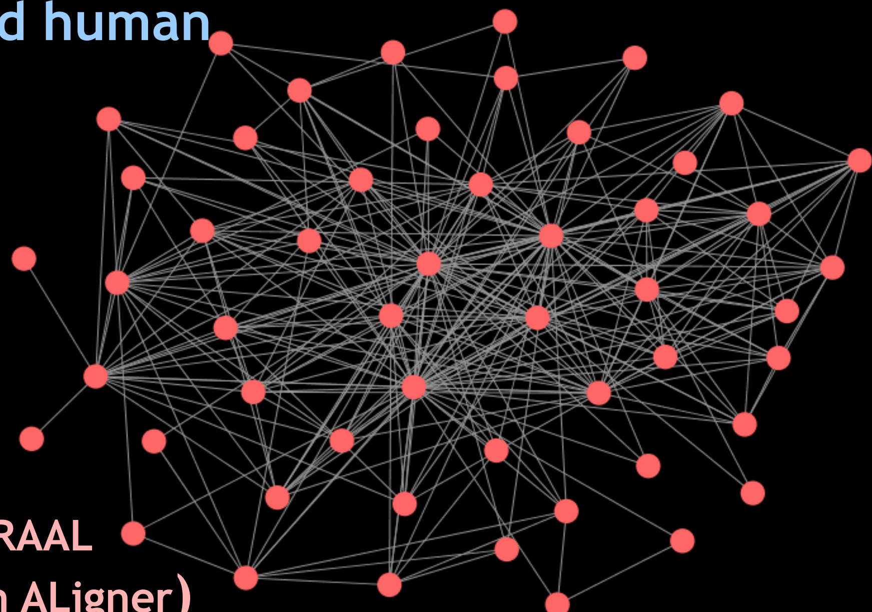
2. Network Analysis and Modeling

Network Alignment

- Find GD node signatures across different networks
- Align "signature-similar" nodes – "seed nodes" in each network
- Extend around seed nodes
  - Greedy seed-and-extend approach
  - Align PPI networks of yeast and human



Isorank



GRAAL
(GRAph ALigner)

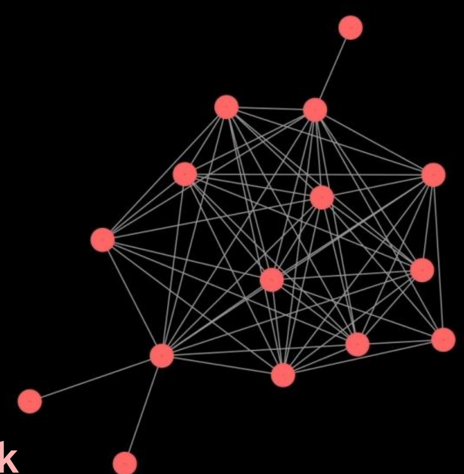
O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes and **N. Przulj**, "Topological Network Alignment Uncovers Biological Function and Phylogeny", *J. Roy Soc. Interface*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

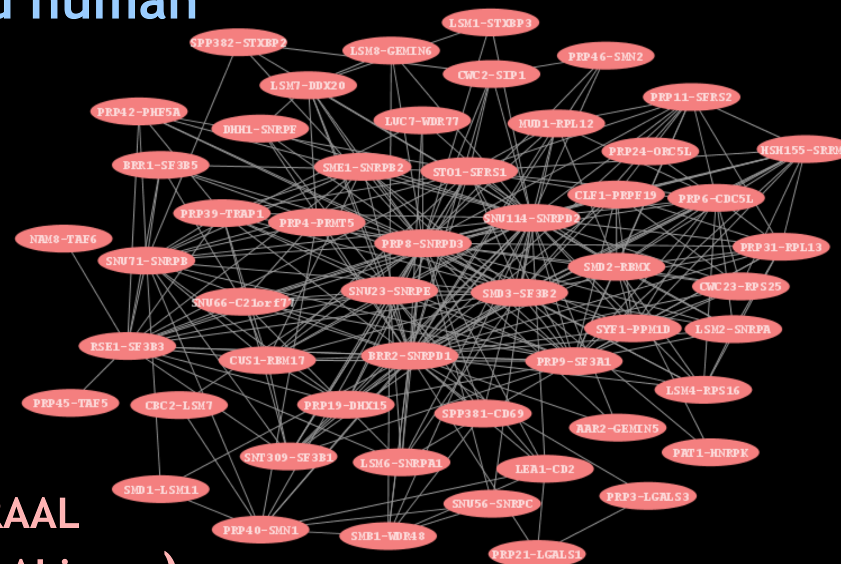
2. Network Analysis and Modeling

Network Alignment

- Find GD node signatures across different networks
- Align "signature-similar" nodes – "seed nodes" in each network
- Extend around seed nodes
  - ➢ Greedy seed-and-extend approach
  - ➢ Align PPI networks of yeast and human



Isorank



GRAAL (GRAph ALigner)

O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes and **N. Przulj**, "Topological Network Alignment Uncovers Biological Function and Phylogeny", *J. Roy Soc. Interface*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Alignment

- Find **GD node signatures** across different networks
- Align "signature-similar" nodes – "**seed nodes**" in each network
- **Extend** around seed nodes
  - Greedy seed-and-extend approach
  - Align **metabolic networks** of Protists

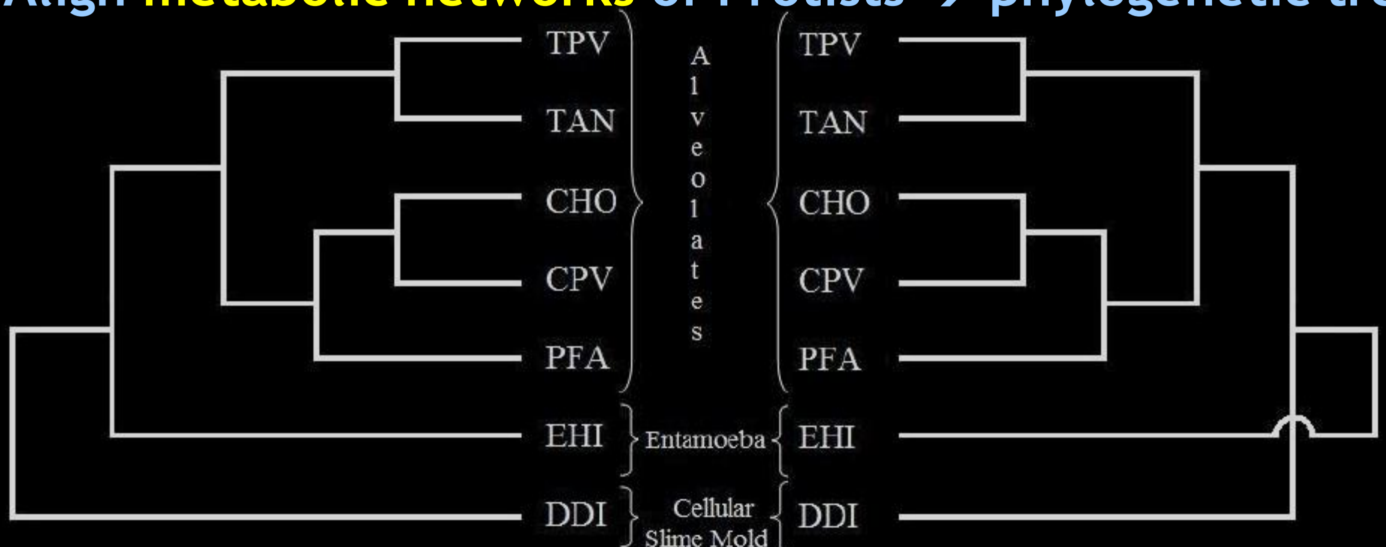
O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes and **N. Przulj**, "Topological Network Alignment Uncovers Biological Function and Phylogeny", *J. Roy Soc. Interface*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Alignment

- Find GD node signatures across different networks
- Align "signature-similar" nodes – "seed nodes" in each network
- Extend around seed nodes
  - Greedy seed-and-extend approach
  - Align metabolic networks of Protists → phylogenetic tree



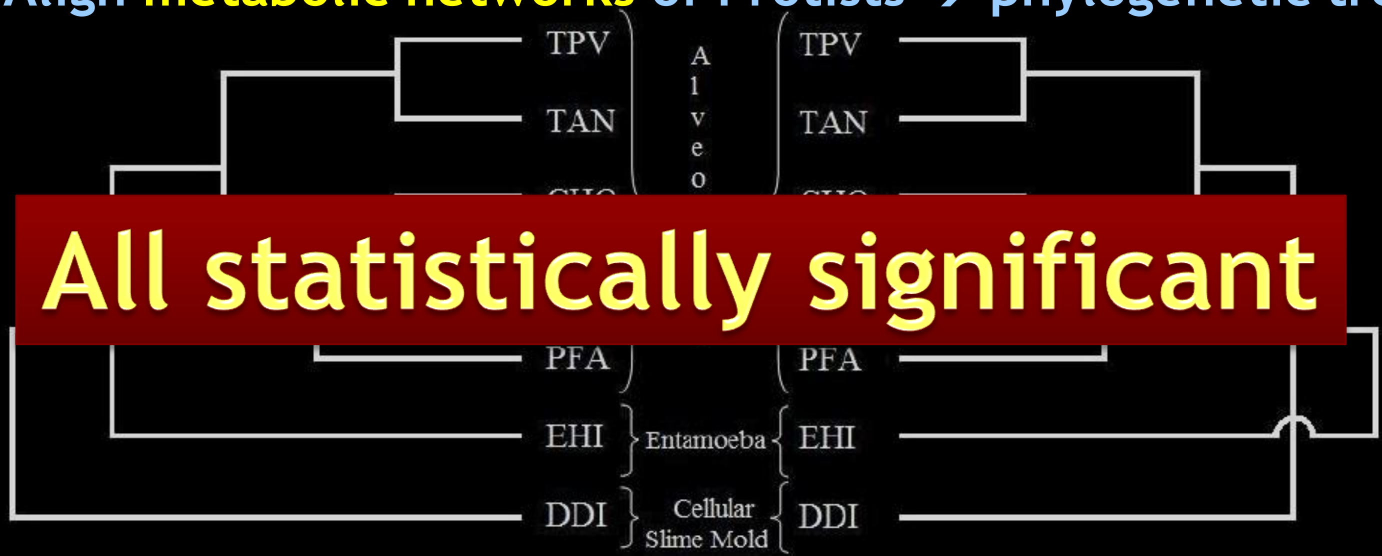
O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes and **N. Przulj**, "Topological Network Alignment Uncovers Biological Function and Phylogeny", *J. Roy Soc. Interface*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Alignment

- Find GD node signatures across different networks
- Align "signature-similar" nodes – "seed nodes" in each network
- Extend around seed nodes
  - ➢ Greedy seed-and-extend approach
  - ➢ Align metabolic networks of Protists → phylogenetic tree



O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes and **N. Przulj**, "Topological Network Alignment Uncovers Biological Function and Phylogeny", *J. Roy Soc. Interface*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Alignment

- **Find GD node signatures across different networks**
- **Align "signature-similar" nodes**
- BUT, not in a seed-and-extend greedy way
- Use the *Hungarian Algorithm* for minimum weight bipartite matching
  - Hence termed ***H-GRAAL***
  - Find ***an optimal*** alignment with respect to the cost function (GDV)
  - ***"Core (stable) alignment"*** - present in all optimal alignments

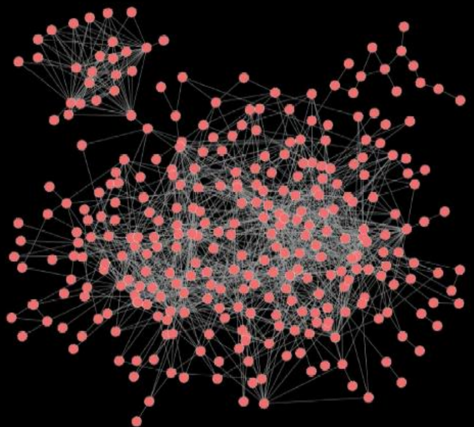
T. Milenkovic, W. L. Ng, W. Hayes and **N. Przulj,** "Optimal Network Alignment Using Graphlet Degree Vectors", *Cancer Informatics*, 9:121-137, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Alignment

- **Find GD node signatures across different networks**
- **Align "signature-similar" nodes**
- BUT, not in a seed-and-extend greedy way
- Use the *Hungarian Algorithm* for minimum weight bipartite matching
  - Hence termed ***H-GRAAL***
  - Find ***an optimal* alignment** with respect to the cost function (GDV)
  - ***"Core (stable) alignment"*** - present in all optimal alignments



Yeast - Human alignment:

- **72% of possible pairs are in core alignment**
- **Both GRAAL & H-GRAAL align: 67% human prots**
- **H-GRAAL core align. in GRAAL: 63% human prots**

T. Milenkovic, W. L. Ng, W. Hayes and **N. Przulj**, "Optimal Network Alignment Using Graphlet Degree Vectors", *Cancer Informatics*, 9:121-137, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

Overview

1. Introduction and Background:
   - PPI and other networks
   - Network analysis and modeling challenges
2. New Network Analysis and Modeling:
   (A) Analysis:
   - ➢ Structure *vs.* biological function
   - ➢ Network alignment

   (B) Modeling – Geometric Graphs:
   - ➢ Graphlet Degree Distributions
   - ➢ Network embedding
   - ➢ Trained Geometric Model
   - ➢ Application: De-noising PPI networks

   (C) Why PPI networks might be geometric?
3. Conclusions

Nataša Pržulj
natasha@imperial.ac.uk

Imperial College London

Overview

1. Introduction and Background:
   - PPI and other networks
   - Network analysis and modeling challenges
2. New Network Analysis and Modeling:
   (A) Analysis:
   - ➢ Structure *vs.* biological function
   - ➢ Network alignment

   (B) Modeling – Geometric Graphs:
   - ➢ Graphlet Degree Distributions
   - ➢ Network embedding
   - ➢ Trained Geometric Model
   - ➢ Application: De-noising PPI networks

   (C) Why PPI networks might be geometric?
3. Conclusions

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

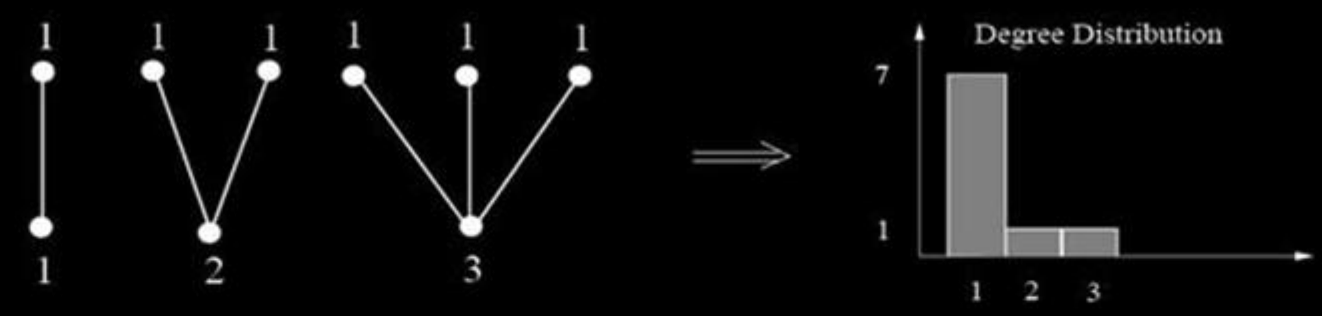
Generalize Degree Distribution

The **degree distribution** measures:

- the number of nodes "touching" k edges for each value of k.

Example:

7 nodes touch 1 edge, 1 touches 2 edges, 1 touches 3 edges.



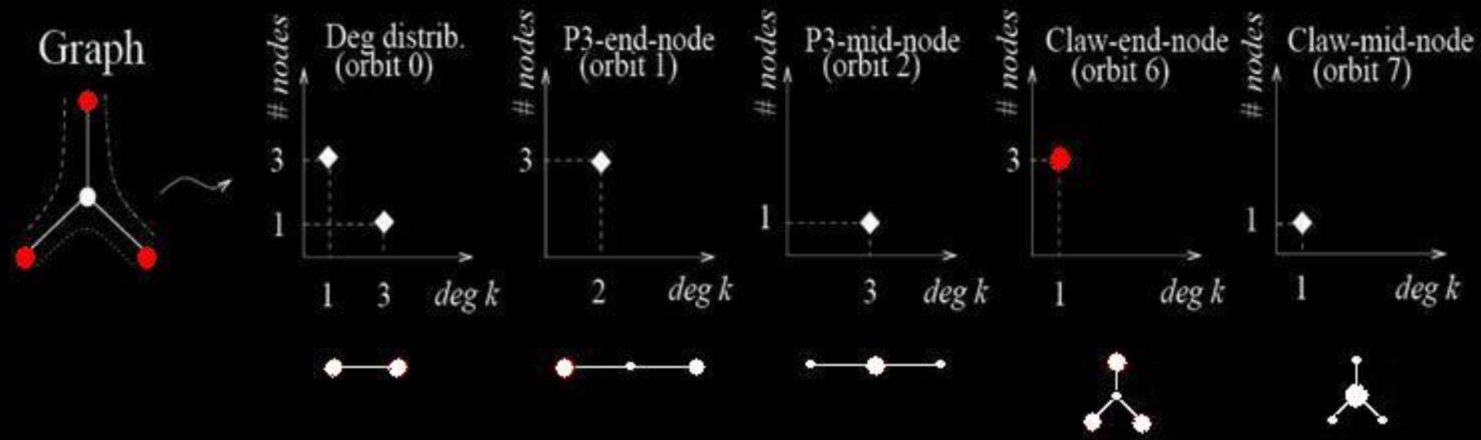
N. Przulj, "Biological Network Comparison Using Graphlet Degree Distribution," *Bioinformatics*, vol. 23, pg. e177-e183, 2007.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

For each of these 73 automorphism orbits, we count:

- the number of nodes touching a particular graphlet at a particular orbit.



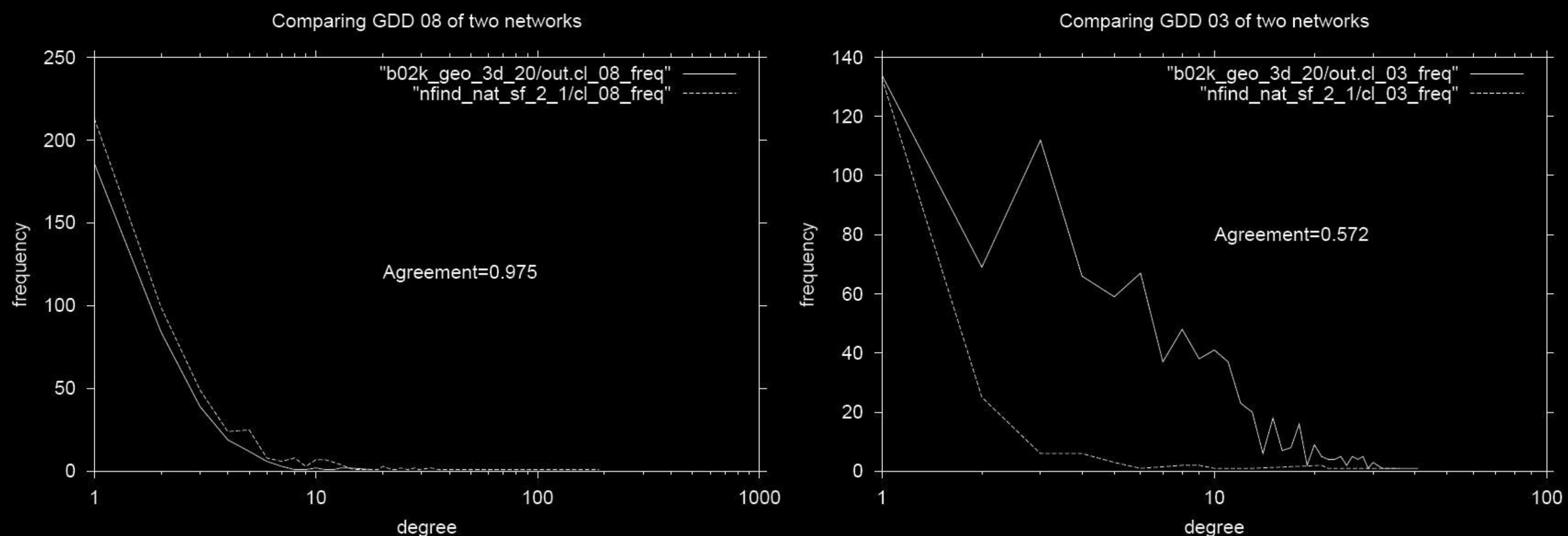
$\Rightarrow$ The *spectrum* of 73 *"graphlet degree distributions (GDDs)"* for $G_0, \ldots, G_{29}$ measuring local structural properties of a network.

N. Przulj, "Biological Network Comparison Using Graphlet Degree Distribution," *Bioinformatics*, vol. 23, pg. e177-e183, 2007.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Illustration of comparing i^{th} GDDs of two networks:



There are 73 such GDD agreements, one for each graphlet orbit.

The *total agreement* between two networks is the average of these 73 GDD agreements.

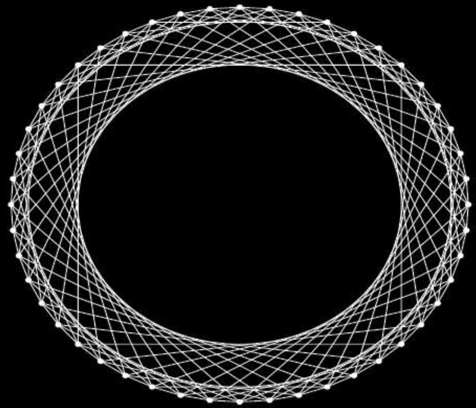
N. Przulj, "Biological Network Comparison Using Graphlet Degree Distribution," *Bioinformatics*, vol. 23, pg. e177-e183, 2007.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Range of Network "Agreement" Measure:

- Agreement between the same model networks: 0.84 ± 0.07
- $\Rightarrow$ Agreements between data and model networks of 0.75 are exceptionally high
- Agreements can be very close to zero:
  e.g. between a $\pm 6, 12$-chord *circulant* and the data < 0.08



N. Przulj, "Biological Network Comparison Using Graphlet Degree Distribution," *Bioinformatics*, vol. 23, pg. e177-e183, 2007.

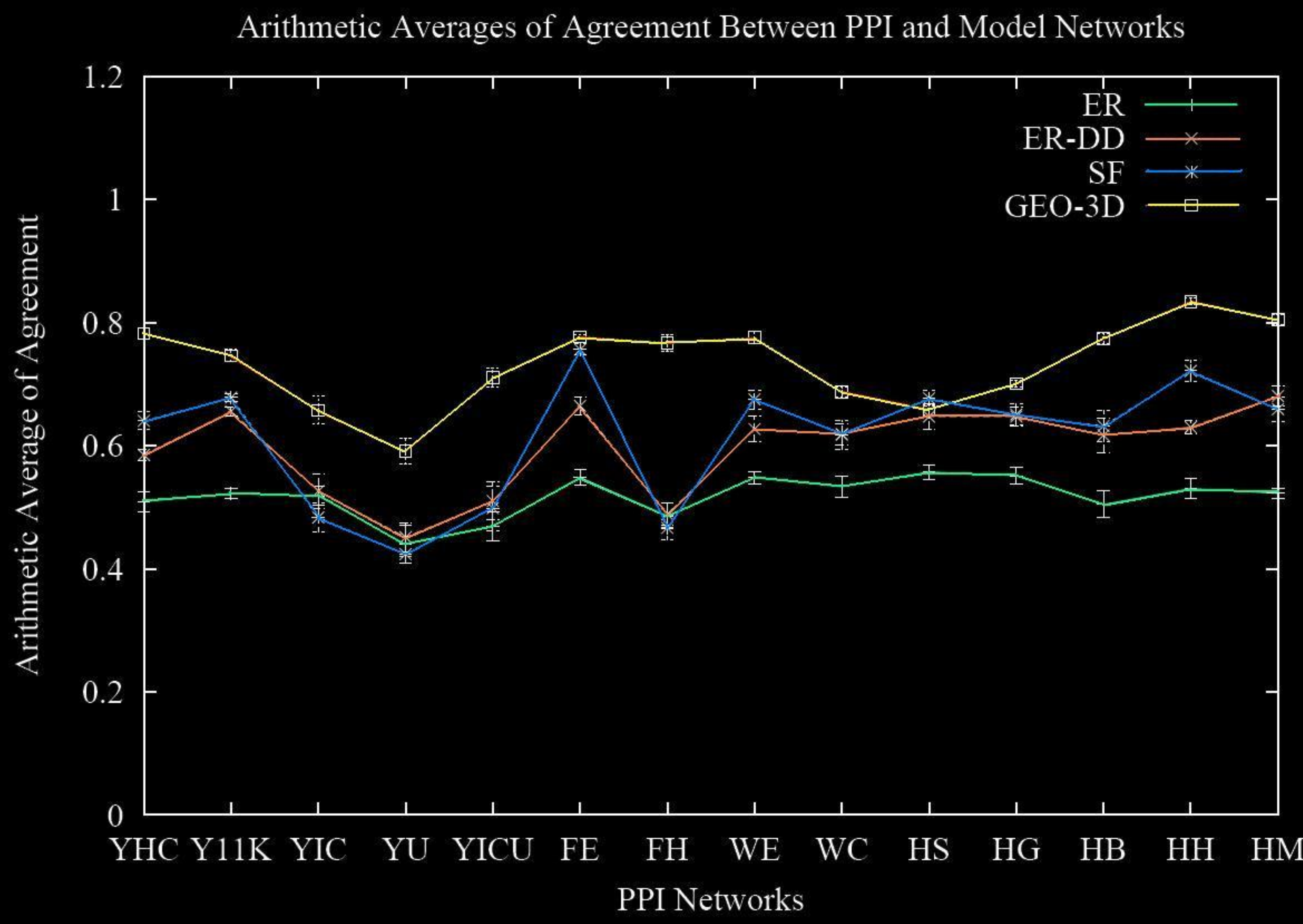
Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

- Yeast *S. cerevisiae*:
  - von Mering (*Nat* 417, 2002): "YHC", "Y11K" (TAP, HMS-PCI)
  - Ito *et al.* (*PNAS* 97, 2000) "core": "YIC" (Y2H)
  - Uetz *et al.* (*Nat* 403, 2000): "YU" (Y2H)
  - Union of "YIC" and "'YU": "YICU" (to increase coverage) (Y2H)
- Fruit fly *D. melanogaster*:
  - Giot *et al.* (*Sci* 302 2003): "FE" and "FH" (Y2H)
- Worm *C. elegans*:
  - Li *et al.* (*Sci* 303, 2004): "WE" and "WC" (Y2H)
- Human:
  - Stelzl (*Cell* 122, 2005): "HS"; Rual (*Nat* 437, 2005): "HG" (Y2H)
  - BIND, HPRD, and MINT: "HB", "HH", "HM" (curated)

N. Przulj, "Biological Network Comparison Using Graphlet Degree Distribution," *Bioinformatics*, vol. 23, pg. e177-e183, 2007.

2. Network Analysis and Modeling



N. Przulj, "Biological Network Comparison Using Graphlet Degree Distribution," *Bioinformatics*, vol. 23, pg. e177-e183, 2007.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

- Different models best-fitting w.r.t. different network properties
- Integrate a variety of local and global network properties into the "network fingerprint" vector:
  1. The average degree
  2. The average clustering coefficient
  3. The average diameter
  4. Frequencies of appearance of all 31 1-5-node graphlets
- Apply a series of machine learning classifiers to network fingerprints:
  1. Backpropagation method (BP)
  2. Probabilistic neural networks (PNN)
  3. Decision tree (DT)
  4. Multinomial naive Bayes classifier (MNB)
  5. Support vector machine (SVM)
- Conclusion: the structure of PPI networks is that of a noisy GEO.

V. Memisevic, T. Milenkovic and **N. Przulj,** "An integrative approach to modeling biological networks", *J. Integrative Bioinformatics*, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

Overview

1. Introduction and Background:
   - PPI and other networks
   - Network analysis and modeling challenges
2. New Network Analysis and Modeling:
   (A) Analysis:
   - Structure *vs.* biological function
   - Network alignment

   (B) Modeling – Geometric Graphs:
   - Graphlet Degree Distributions
   - Network embedding
   - Trained Geometric Model
   - Application: De-noising PPI networks

   (C) Why PPI networks might be geometric?
3. Conclusions

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Embedding Algorithm

- Use Multi Dimensional Scaling (MDS)
- Paths play a role of distances

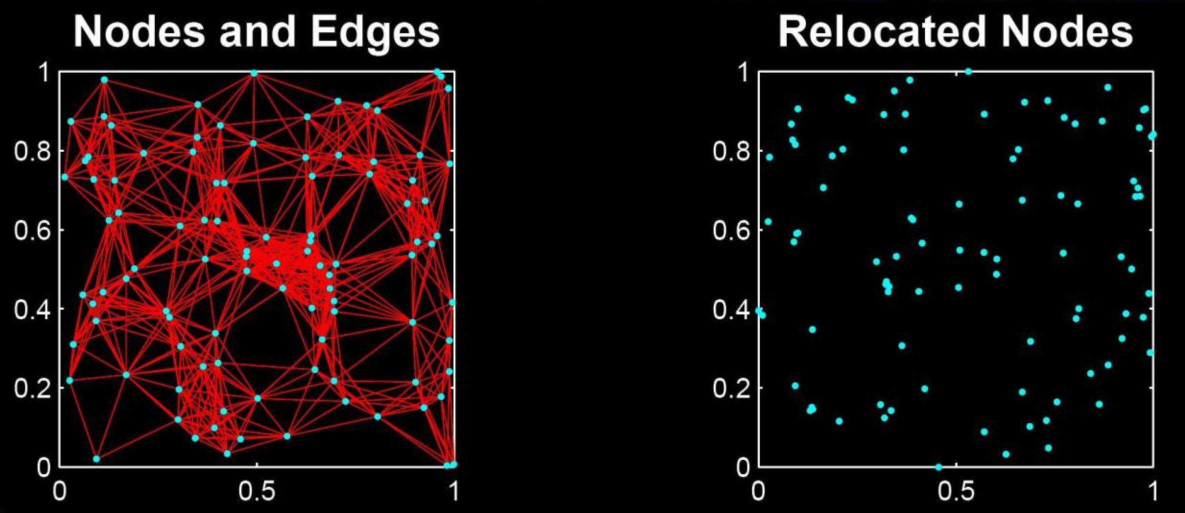
D. J. Higham, M. Rasajski, **N. Przulj**, "Fitting a Geometric Graph to a Protein-Protein Interaction Network", *Bioinformatics*, 24(8), 1093-1099, 2008.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Embedding Algorithm

- Use Multi Dimensional Scaling (MDS)
- Paths play a role of distances



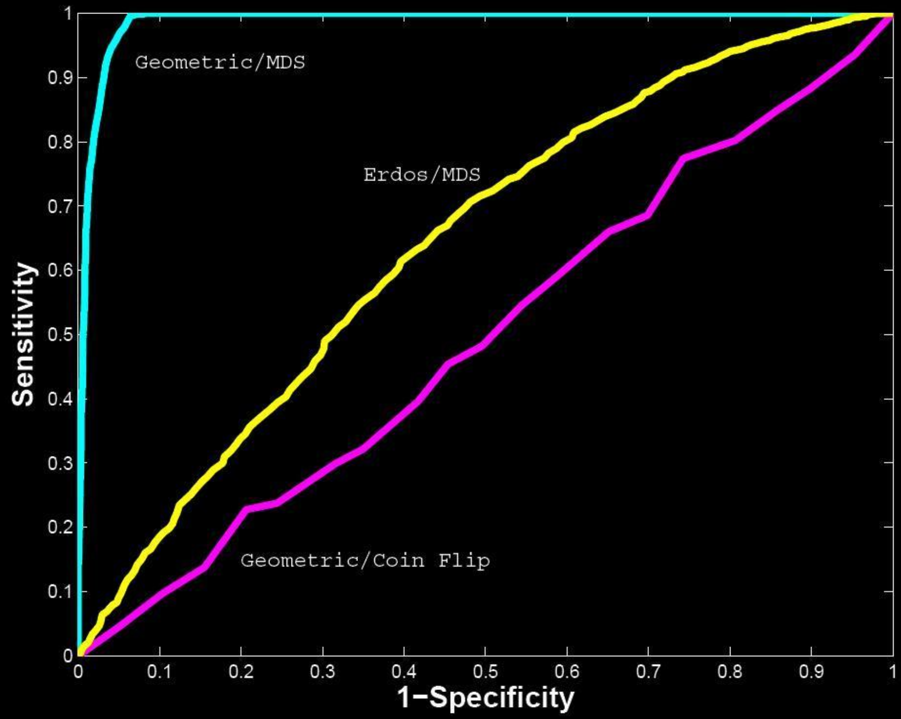
D. J. Higham, M. Rasajski, **N. Przulj**, “Fitting a Geometric Graph to a Protein-Protein Interaction Network”, *Bioinformatics*, 24(8), 1093-1099, 2008.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Embedding Algorithm

- Use Multi Dimensional Scaling (MDS)
- Paths play a role of distances



TN=no edge in original,
no edge in embedded

…

Specificity=TN/(TN+FP)

Sensitivity=TP/(TP+FN)

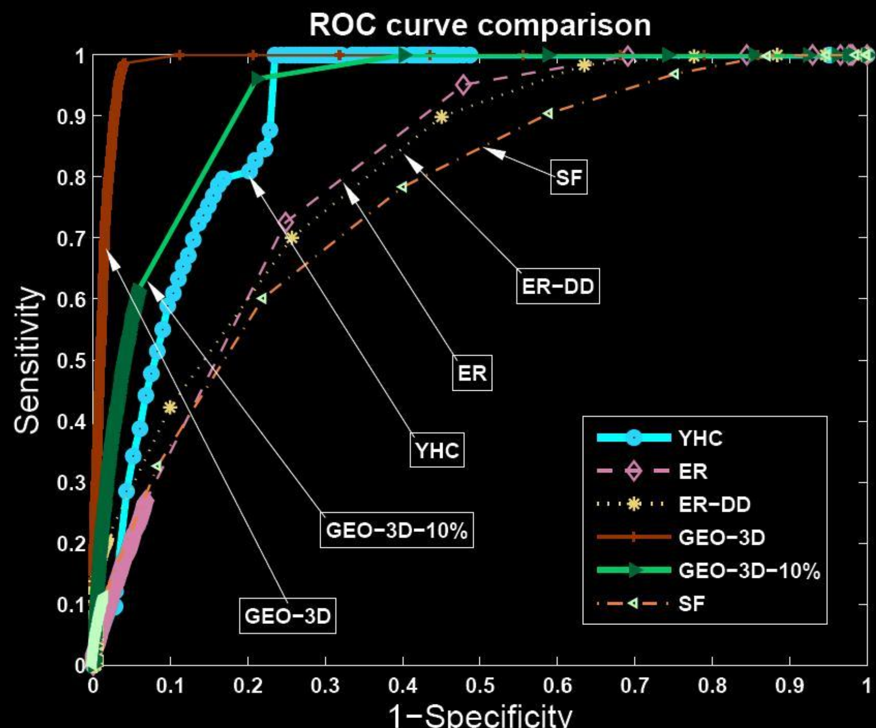
D. J. Higham, M. Rasajski, **N. Przulj**, "Fitting a Geometric Graph to a Protein-Protein Interaction Network", *Bioinformatics*, 24(8), 1093-1099, 2008.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Network Embedding Algorithm

- Use Multi Dimensional Scaling (MDS);
- Paths play a role of distances



D. J. Higham, M. Rasajski, **N. Przulj**, "Fitting a Geometric Graph to a Protein-Protein Interaction Network", *Bioinformatics*, 24(8), 1093-1099, 2008.

Nataša Pržulj
natasha@imperial.ac.uk

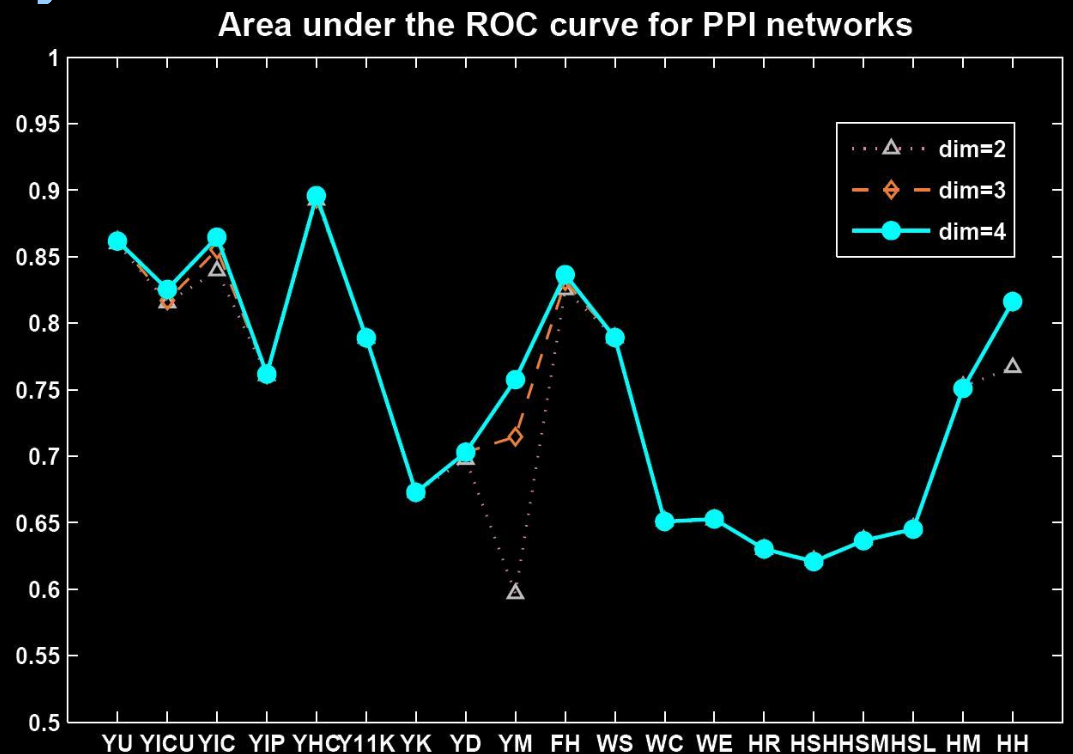
2. Network Analysis and Modeling

Network Embedding Algorithm

- Use Multi Dimensional Scaling (MDS)
- Paths play a role of distances

1.0 →

0.5 →



D. J. Higham, M. Rasajski, **N. Przulj**, "Fitting a Geometric Graph to a Protein-Protein Interaction Network", *Bioinformatics*, 24(8), 1093-1099, 2008.

Nataša Pržulj
natasha@imperial.ac.uk

Overview

1. Introduction and Background:
   - PPI and other networks
   - Network analysis and modeling challenges
2. New Network Analysis and Modeling:
   - (A) Analysis:
     - Structure *vs.* biological function
     - Network alignment
   - (B) Modeling – Geometric Graphs:
     - Graphlet Degree Distributions
     - Network embedding
     - Trained Geometric Model
     - Application: De-noising PPI networks
   - (C) Why PPI networks might be geometric?
3. Conclusions

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

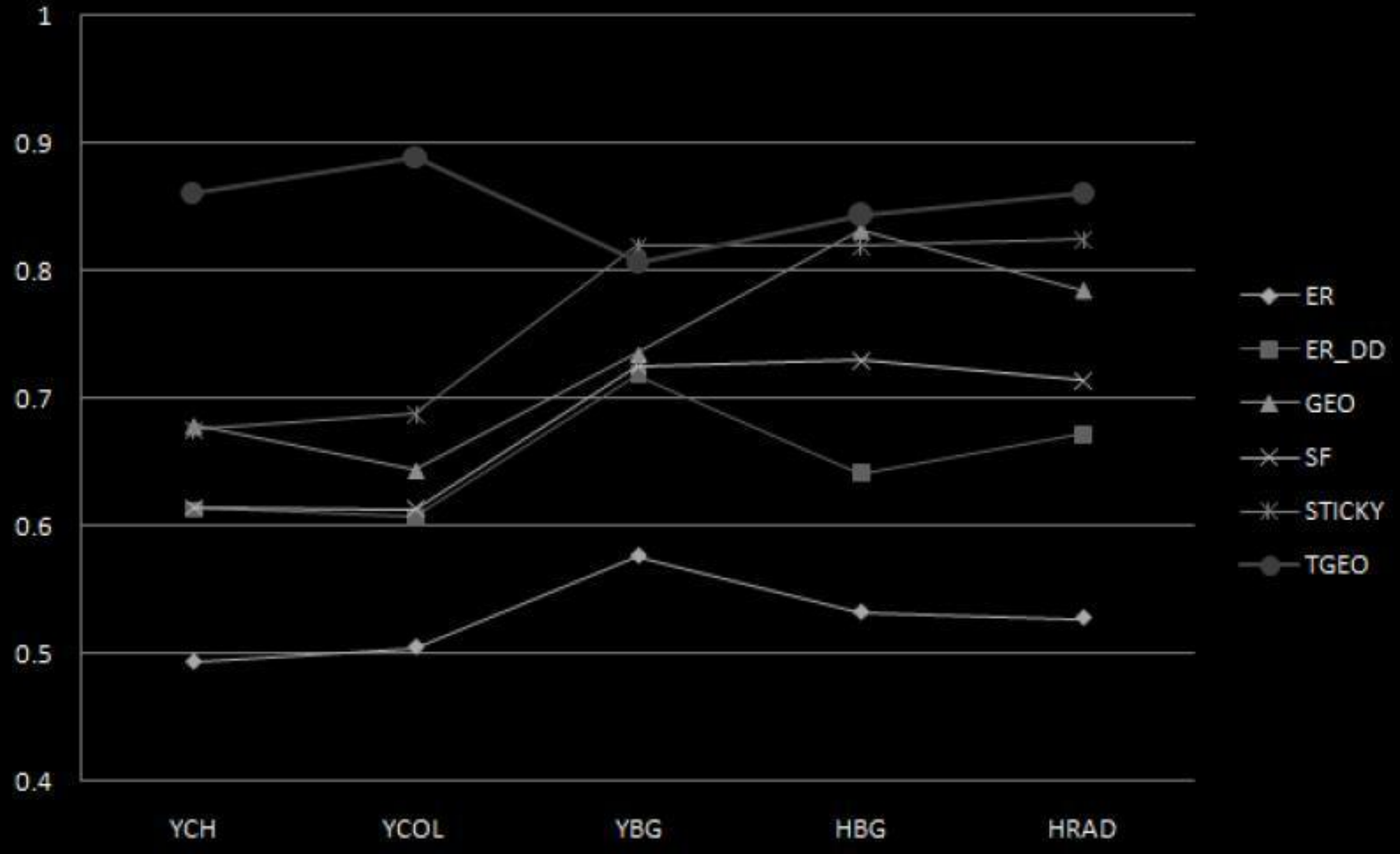
Trained Geometric Model

- Embed a high-confidence PPI network in space
- Learn distribution of points in this space using:
  - Mixture of Gaussians
  - Learn the required parameters using Expectation-Maximization algorithm (EM)
- Generate model networks of arbitrary size by:
  - distributing points in the space according to the trained distribution
  - connecting two nodes by an edge if they are close enough in space.

O. Kuchaiev and **N. Przulj**, "Learning the Structure of Protein-Protein Interaction Networks", *Pacific Symposium on Biocomputing (PSB'09)*, Hawaii, USA, 2009.

2. Network Analysis and Modeling

Trained Geometric Model



GDD-agreement between the data and model networks.

O. Kuchaiev and **N. Przulj**, "Learning the Structure of Protein-Protein Interaction Networks", *Pacific Symposium on Biocomputing (PSB'09)*, Hawaii, USA, 2009.

Nataša Pržulj
natasha@imperial.ac.uk

Overview

1. Introduction and Background:
   - PPI and other networks
   - Network analysis and modeling challenges
2. New Network Analysis and Modeling:

   (A) Analysis:
   - ➢ Structure *vs.* biological function
   - ➢ Network alignment

   (B) Modeling – Geometric Graphs:
   - ➢ Graphlet Degree Distributions
   - ➢ Network embedding
   - ➢ Trained Geometric Model
   - ➢ Application: De-noising PPI networks

   (C) Why PPI networks might be geometric?
3. Conclusions

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Application: De-Noising PPI Networks

- Embed a PPI network into space
- Learn from cooridates of node embedding points probability densities *p(dist|edge)* and *p(dist|non-edge)*:
  - ➢ Mixture of Gaussians
  - ➢ Learn parameters using Expectation-Maximization (EM)

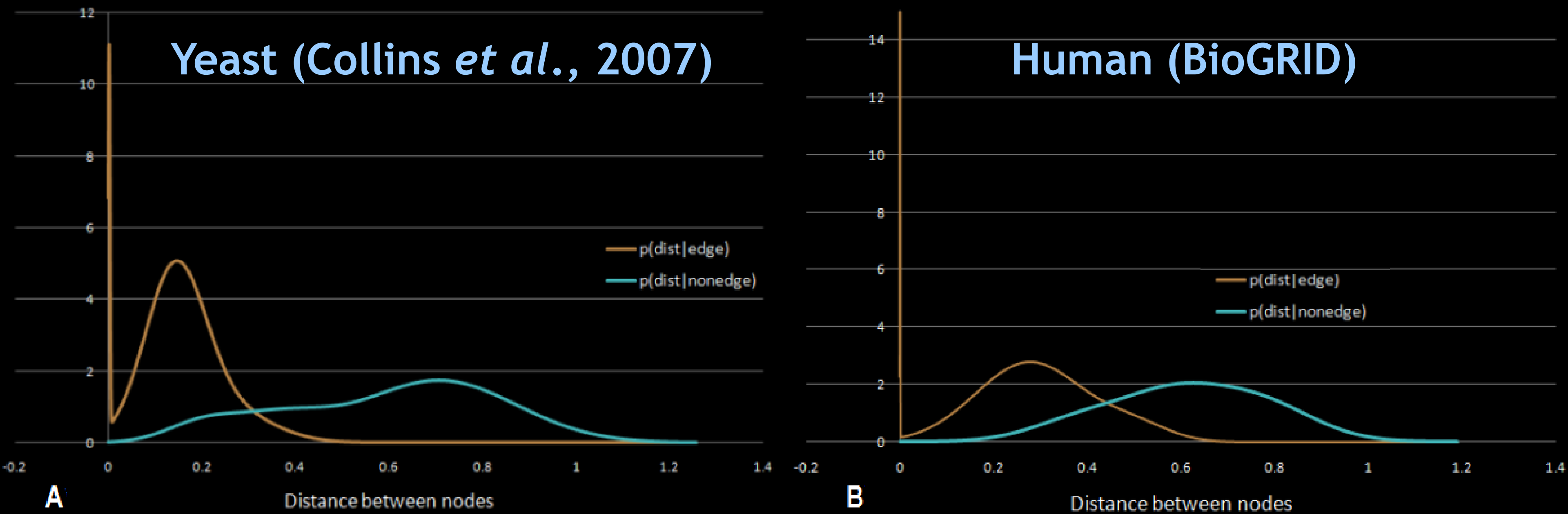
O. Kuchaiev, M. Rasajski, D. Higham, and **N. Przulj**, "Geometric De-noising of Protein-Protein Interaction Networks", *PLoS Computational Biology* 5(8), e1000454, 2009.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Application: De-Noising PPI Networks

- Embed a PPI network into space
- Learn from cooridates of node embedding points probability densities *p(dist|edge)* and *p(dist|non-edge)*:



O. Kuchaiev, M. Rasajski, D. Higham, and **N. Przulj**, “Geometric De-noising of Protein-Protein Interaction Networks”, *PLoS Computational Biology* 5(8), e1000454, 2009.

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Application: De-Noising PPI Networks

- Embed a PPI network into space
- Learn from cooridates of node embedding points probability densities *p(dist|edge)* and *p(dist|non-edge)*
- Choose a threshold δ
- For each pair of nodes at distance ≤ δ, compute its Confidence Score:

$$CS(i,j) = \frac{p(edge(i,j)|dist(i,j))}{p(edge(i,j)|dist(i,j)) + p(nonedge(i,j)|dist(i,j))}$$

⇒ Predict new PPIs

O. Kuchaiev, M. Rasajski, D. Higham, and **N. Przulj**, "Geometric De-noising of Protein-Protein Interaction Networks", *PLoS Computational Biology* 5(8), e1000454, 2009.

Nataša Pržulj
natasha@imperial.ac.uk

Imperial College
London

Overview

1. Introduction and Background:
   - PPI and other networks
   - Network analysis and modeling challenges
2. New Network Analysis and Modeling:
   (A) Analysis:
   - ➢ Structure *vs.* biological function
   - ➢ Network alignment

   (B) Modeling – Geometric Graphs:
   - ➢ Graphlet Degree Distributions
   - ➢ Network embedding
   - ➢ Trained Geometric Model
   - ➢ Application: De-noising PPI networks

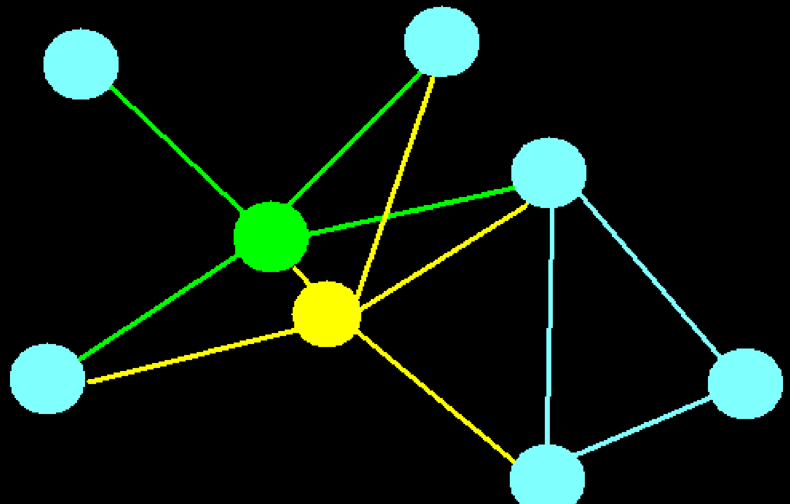
   (C) Why PPI networks might be geometric?
3. Conclusions

Nataša Pržulj
natasha@imperial.ac.uk

2. Network Analysis and Modeling

Why PPI networks might be geometric?

- Intuitive “geometricity” of PPI networks:
  - Genes exist in some bio-chemical space
  - Gene duplications and mutations
  - Natural selection = “evolutionary optimization”



N. Przulj, O. Kuchaiev, A. Stevanovic, and W. Hayes “Geometric Evolutionary Dynamics of Protein Interaction Network”, *Pacific Symposium on Biocomputing (PSB’10)*, Hawaii, 2010.

Nataša Pržulj
natasha@imperial.ac.uk

Overview

1. Introduction and Background:
   - PPI and other networks
   - Network analysis and modeling challenges
2. New Network Analysis and Modeling:
   (A) Analysis:
   - ➢ Structure *vs.* biological function
   - ➢ Network alignment

   (B) Modeling – Geometric Graphs:
   - ➢ Graphlet Degree Distributions
   - ➢ Network embedding
   - ➢ Trained Geometric Model
   - ➢ Application: De-noising PPI networks

   (C) Why PPI networks might be geometric?
3. Conclusions

Nataša Pržulj
natasha@imperial.ac.uk

Conclusions

New Network Analysis and Modeling:

(A) Analysis – Graphlet Signatures :

- Structure *vs.* biological function and disease
- Homology from topology
- Network alignment

(B) Modeling – Geometric Graphs:

- Graphlet Degree Distributions
- Network embedding
- Trained Geometric Model
- Application: De-noising PPI networks

(C) Why PPI networks might be geometric?

Nataša Pržulj
natasha@imperial.ac.uk

Other Research Projects

1) *GraphCrunch* - network analysis software:
   - Milenković, Lai, & Pržulj, *BMC Bioinformatics*, 2008. Highly Accessed.
2) Residue Interaction Graphs:

   \- New null model:
   - Milenković, Filippis, Lappe, & Pržulj, *PLoS ONE* 4(6), 2009.
3) Structure of brain functional networks:
   - Kuchaiev, Weng, Nenadić, & Przulj, *IEEE Engineering in Medicine and Biology Society (EMBC'09)*, ***2009***.
4) Planar Cell Polarity - ommatidial rotation in fruit-fly eye caused by net. structural perturbation

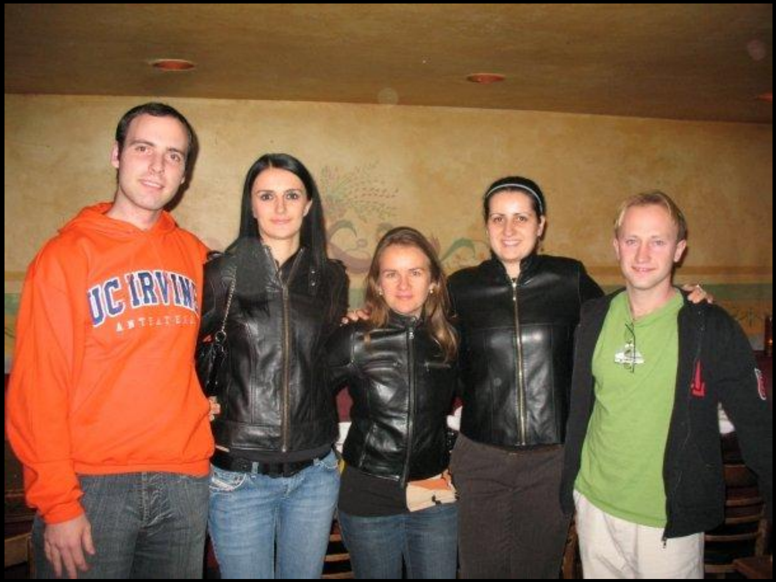
5) **Etc.**

Nataša Pržulj
natasha@imperial.ac.uk

Acknowledgements

- <u>Funding</u>: **NSF CAREER** (Jan. 2007 — Dec. 2011)
  NSF CDI, $2M (Oct. 2010 – 2014)
  Center for Complex Bio. Sys. at UCI (2008—2009)
  Council on Research & Computing UCI (2008—2009)
- <u>Alumni</u>:
  1. Tijana Milenković, Ph.D.
     Assistant Prof., U. of Notre Dame
  2. Oleksii Kuchaiev, Ph.D.
     Microsoft, Redmond
  3. Vesna Memišević, Ph.D.
     US Army, Bioinformatics Res.
  4. Aleksandar Stevanović, M.Sc.

  Undergraduate programmer: Jason Lai
- <u>Collaborators</u>: D.Higham, L.Huang, M.Lappe,
  W.Hayes, Z.Nenadić, P.Kaiser, A.Ganesan…



1. O. Kuchaiev, T. Milenkovic, V. Memisevic, W. Hayes and **N. Przulj**, "Topological Network Alignment Uncovers Biological Function and Phylogeny", *J. Roy Soc. Interface*, 2010.
2. **N. Przulj**, O. Kuchaiev, A. Stevanovic, and W. Hayes "Geometric Evolutionary Dynamics of Protein Interaction Network", *Pacific Symposium on Biocomputing (PSB'10)*, Hawaii, USA, 2010.
3. T. Milenkovic, W. L. Ng, W. Hayes and **N. Przulj**, "Optimal Network Alignment Using Graphlet Degree Vectors", *Cancer Informatics*, 9:121-137, 2010.
4. **V. Memisevic**, T. Milenkovic and **N. Przulj**, "Complementarity of network and sequence information in homologous proteins", *J. Integrative Bioinformatics*, 2010.
5. T. Milenkovic, V. Memisevic, A. K. Ganesan, and **N. Przulj**, "Systems-level Cancer Gene Identification from Protein Interaction Network Topology Applied to Melanogenesis-related Interaction Networks", *J. Roy. Soc. Interface*, 2009.
6. **O. Kuchaiev, M. Rasajski, D. Higham**, and **N. Przulj**, "Geometric De-noising of Protein-Protein Interaction Networks", *PLoS Computational Biology* 5(8), e1000454, 2009.
7. **O. Kuchaiev** and **N. Przulj**, "Learning the Structure of Protein-Protein Interaction Networks", *Pacific Symposium on Biocomputing (PSB'09)*, Hawaii, USA, 2009.
8. C. Guerrero, T. Milenkovic, **N. Przulj**, P. Keiser, L. Huang, "Characterization of the proteasome interaction network using a QTAX-based tag-team strategy and protein interaction network analysis," *PNAS*, 105 (36), pg. 13333-13338 2008.
9. T. Milenkovic and **N. Przulj**, "Uncovering Biological Network Function via Graphlet Degree Signatures," *Cancer Informatics*, vol. 4, pg. 257-273, 2008.
10. D. J. Higham, M. Rasajski, **N. Przulj**, "Fitting a Geometric Graph to a Protein-Protein Interaction Network," *Bioinformatics*, vol. 24, num. 8, pg. 1093-1099, 2008.
11. T. Milenkovic, J. Lai, **N. Przulj**, "GraphCrunch: A Tool for Large Network Analyses," *BMC Bioinformatics*, 9:70, January 30, 2008. Highly Accessed.
12. F. Hormozdiari, P. Berenbrink, **N. Przulj**, C. Sahinalp, "Not all Scale Free Networks are Born Equal: the Role of the Seed Graph in PPI Network Emulation," *PLoS Computational Biology*, 3(7), 2007.
13. **N. Przulj**, "Geometric Local Structure in Biological Networks," *IEEE ITW'07 Invited Paper*, 2007.
14. **N. Przulj**, "Biological Network Comparison Using Graphlet Degree Distribution," *Bioinformatics* proc. of *ECCB'06*,23:e177-e183, 2007.
15. **N. Przulj** and D. Higham, "Modelling Protein-Protein Interaction Networks via a Stickiness Index," *J. of the Royal Society Interface*, vol. 3, num. 10, pg 711 - 716, 2006.
16. **N. Przulj**, D. G. Corneil, and I. Jurisica, "Efficient Estimation of Graphlet Frequency Distributions in Protein-Protein Interaction Networks," *Bioinformatics*, vol. 22, num. 8, pg 974-980, 2006.
17. M. Barrios-Rodiles, K. R. Brown, B. Ozdamar, Z. Liu, R. S. Donovan, F. Shinjo, Y. Liu, R. Bose, J. Dembowy, I. W. Taylor, V. Luga, **N. Przulj**, M. Robinson, H. Suzuki, Y. Hayashizaki, I. Jurisica, and J. L. Wrana, "High-Throughput Mapping of a Dynamic Signaling Network in Mammalian Cells," ***Science***, vol. 307, num. 5715, pg. 1621-1625, 2005.
18. **N. Przulj**, D. G. Corneil, and I. Jurisica, "Modeling Interactome: Scale-Free or Geometric?," *Bioinformatics*, vol. 20, num. 18, pg. 3508-3515, 2004.
19. **N. Przulj**, D. Wigle, and I. Jurisica, "Functional Topology in a Network of Protein Interactions," *Bioinformatics*, vol. 20, num. 3, pg. 340-348, 2004.